

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

Інститут телекомунікаційних систем

Кафедра Інформаційно-телекомунікаційних мереж

«На правах рукопису»

УДК 004.021

«До захисту допущено»

Завідувач кафедри

_____ Лариса ГЛОБА

« ____ » _____ 2020 р.

**Магістерська дисертація
на здобуття ступеня магістра
за освітньо-професійною програмою «Інформаційно-комунікаційні
технології»
зі спеціальності 172 «Телекомунікації та радіотехніка»
на тему: «Модифікований метод розподілу навантаження в
інфокомунікаційних мережах»**

Виконав:

студент VI курсу, групи ПІ-91мп

Хрищенко Роман Андрійович _____

Керівник:

доцент кафедри ІТМ ІТС, доцент, к.т.н.

Правило Валерій Володимирович _____

Рецензент:

доцент кафедри ТК ІТС, доцент, к.т.н.

Явіся Валерій Сергійович _____

Засвідчую, що у цій магістерській
дисертації немає запозичень з праць
інших авторів без відповідних посилань.

Студент _____

Київ – 2020 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Інститут телекомунікаційних систем
Кафедра Інформаційно-телекомунікаційних мереж

Рівень вищої освіти – другий (магістерський)

Спеціальність – 172 «Телекомунікації та радіотехніка»

Освітньо-професійна програма «Інформаційно-комунікаційні технології»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Лариса ГЛОБА

«__» _____ 2020 р.

ЗАВДАННЯ
на магістерську дисертацію студенту
Хрищенюку Роману Андрійовичу

1. Тема дисертації «Модифікований метод розподілу навантаження в інфокомунікаційних мережах», науковий керівник дисертації доцент кафедри інформаційно-телекомунікаційних мереж ІТС Правило Валерій Володимирович, доцент, к.т.н., затверджені наказом по університету від «03» листопада 2020 р. № 3208-с.
2. Термін подання студентом дисертації 10.12.2020 р.
3. Об'єкт дослідження: Процес розподілу навантаження в інфокомунікаційній мережі.
4. Предмет дослідження: Метод розподілу навантаження в інфокомунікаційній мережі
5. Перелік завдань, які потрібно розробити:
 1. Огляд структури, технологічних компонентів та принципів побудови інфокомунікаційної мережі.
 2. Аналіз основних причин перенавантаження інфокомунікаційної мережі.
 3. Аналіз переваг та недоліків проаналізованих методів розподілу навантаження в ІКМ.

4. Аналіз існуючих методів розподілу навантаження в ІКМ.
 5. Розробка модифікованого методу розподілу навантаження в інфокомунікаційних мережах, на основі проведеного аналізу.
 6. Дослідити можливість застосування запропонованого методу експериментально та шляхом імітаційного моделювання
6. Орієнтовний перелік ілюстративного матеріалу:
1. Тема, мета, актуальність, задачі дослідження
 2. Огляд структури, технологічних компонентів та принципів побудови інфокомунікаційної мережі
 3. Аналіз існуючих методів розподілу навантаження в ІКМ
 4. Розробка модифікованого методу розподілу навантаження в інфокомунікаційних мережах, на основі проведеного аналізу.
 5. Результати експерименту та імітаційного моделювання роботи запропонованого підходу
 6. Опис запропонованого підходу як стартап-проекту
 7. Загальні висновки.
7. Орієнтовний перелік публікацій:
8. Дата видачі завдання 28.10.2019р.
-

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Ознайомлення з технологічними особливостями ІКМ	01.09.2019 – 30.09.2020	виконано
2	Огляд принципів побудови та роботи ІКМ	01.10.2019 – 30.11.2020	виконано
3	Виділення основних причин перенавантаження ІКМ	01.12.2019 – 28.02.2020	виконано
4	Аналіз існуючих методів розподілу навантаження в ІКМ	01.03.2019 – 15.04.2020	виконано
5	Формування порівняльної характеристики проаналізованих методів розподілу навантаження	16.04.2019 – 31.05.2020	виконано
6	Розробка модифікованого методу раціонального розподілу навантаження в ІКМ	1.06.2019 – 30.06.2020	виконано

7	Дослідження застосування запропонованого методу експериментально та шляхом імітаційного моделювання	1.07.2019 - 30.09.2020	виконано
---	---	------------------------	----------

Студент

Роман ХРИЩЕНЮК

Науковий керівник дисертації

Валерій ПРАВИЛО

РЕФЕРАТ

Робота містить 95 сторінок, 41 рисунок та 26 таблиць. Було використано 30 джерел.

Актуальність теми полягає в тому, що в зв'язку з швидким накопиченням об'ємів даних та постійним перенавантаженням інфокомунікаційних мереж, постає потреба у створенні модифікованого методу розподілу навантаження для покращення роботи інфокомунікаційної мережі.

Мета дослідження: забезпечення раціонального розподілу навантаження в інфокомунікаційній мережі за рахунок використання запропонованого методу.

Задачі дослідження:

1. Огляд структури, технологічних компонентів та принципів побудови інфокомунікаційної мережі.

2. Аналіз основних причин перенавантаження інфокомунікаційної мережі.

3. Аналіз існуючих методів розподілу навантаження в ІКМ

4. Виділення переваг та недоліків проаналізованих методів розподілу навантаження в ІКМ.

5. Розробка модифікованого методу розподілу навантаження в інфокомунікаційних мережах, на основі проведеного аналізу.

6. Дослідити можливість застосування запропонованого методу експериментально та шляхом імітаційного моделювання.

Об'єкт дослідження: процес розподілу навантаження в інфокомунікаційній мережі.

Предмет дослідження: метод розподілу навантаження в інфокомунікаційній мережі.

Методи дослідження: основними методами дослідження є математичне та імітаційне моделювання.

Наукова новизна одержаних результатів:

На основі проведеного аналізу був розроблений модифікований метод з прогнозуванням планування розподілу навантаження в інфокомунікаційних

мережах на основі циклічного планування Round Robin. Новизна методу полягає в прогнозуванні поточного стану завантаження веб-серверів в кластері, а також в можливості отримання робочого стану, споживання ємності та допустимого навантаження веб-серверів і взагалі продуктивності всього веб-кластеру, що дозволяє підвищити відмовостійкість та покращити ефективність роботи інфокомунікаційної мережі.

Практичне значення одержаних результатів:

Результати дослідження дозволять підвищити ефективність роботи інфокомунікаційної мережі, за допомогою використання запропонованого методу розподілу навантаження в мережі. Проведений аналіз та експериментальне дослідження запропонованого методу.

Ключові слова: інфокомунікаційна мережа, розподіл навантаження, round robin.

ABSTRACT

The work contains 95 pages, 41 figures and 26 tables. 30 sources were used.

Relevance of the topic:

The relevance of the topic is that due to the rapid accumulation of data and the constant overload of infocommunication networks, there is a need to create a modified method of load distribution to improve the performance of the infocommunication network.

The purpose of study is to ensure the rational distribution of load in the infocommunication network through the use of the proposed method.

Research objectives:

1. Review of the structure, technological components and principles of construction of the infocommunication network.
2. Analysis of the main causes of overload in the infocommunication network.
3. Analysis of existing methods of load distribution in the infocommunication network.
4. Formation of advantages and disadvantages of the analyzed methods of load distribution in the infocommunication network.
5. Development of a modified method of load distribution in infocommunication networks, based on the analysis.
6. To explore the possibility of the proposed method application experimentally and through simulation

Object of research: the process of load distribution in the infocommunication network.

Subject of research: method of load distribution in the infocommunication network.

Methods of research: the main methods of research are mathematical modeling and simulation.

Scientific novelty of the obtained results:

Based on the analysis, a modified method with developed to predict load balancing planning in infocommunication networks based on Round Robin cyclic planning. The novelty of the method is to predict the current state of loading the web

server in the cluster, as well as the ability to obtain operating status, capacity consumption and allowable load of web servers and overall performance of the entire web cluster, which increases fault tolerance and improves the efficiency of the infocommunication network.

The practical value of the results obtained:

The results of the study will increase the efficiency of the infocommunication network by using the proposed method of load distribution in the network. The analysis and experimental research of the offered method are carried out.

Keywords: infocommunication network, load distribution, round robin.

ЗМІСТ

ВСТУП.....	12
РОЗДІЛ 1 ЗАГАЛЬНИЙ ОГЛЯД РОБОТИ ІНФОКОМУНІКАЦІЙНОЇ МЕРЕЖІ.....	14
1.1 Загальні відомості та можливості мереж	14
1.2 Структура мережі	21
1.3 Основні технологічні компоненти мережі	24
1.4 Принципи побудови та роботи інфокомунікаційної мережі.....	26
РОЗДІЛ 2 ПЕРЕНАВАНТАЖЕННЯ МЕРЕЖІ ТА АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ РОЗПОДІЛУ НАВАНТАЖЕННЯ.....	36
2.1 ОГЛЯД ТА ВИДІЛЕННЯ ОСНОВНИХ ПРИЧИН ПЕРЕНАВАНТАЖЕННЯ МЕРЕЖІ	36
2.1.1 Причини перенавантаження використовуючи комутацію каналів в мережі.....	39
2.1.2 Причини перенавантаження використовуючи комутацію повідомлень в мережі	41
2.1.3 Причини перенавантаження використовуючи комутацію пакетів в мережі.....	43
2.1.4 Причини перенавантаження використовуючи комутацію комірок в мережі.....	45
2.2 ОСНОВНІ ЦІЛІ РОЗПОДІЛУ НАВАНТАЖЕННЯ В ІНФОКОМУНІКАЦІЙНІЙ МЕРЕЖІ.....	48
2.3 АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ РОЗПОДІЛУ НАВАНТАЖЕННЯ В ІКМ	51
2.3.1 Статичний розподіл навантаження.....	51
2.3.2 Динамічний розподіл навантаження	52

	10
2.3.3 Розподіл навантаження на транспортному рівні.....	53
2.3.4 Розподіл навантаження на прикладному рівні	55
2.3.5 Розподіл навантаження на мережевому рівні.....	57
2.3.6 Round Robin.....	58
2.3.7 Weighted Round Robin	63
2.3.8 Least Connections.....	64
2.4 ПЕРЕВАГИ ТА НЕДОЛІКИ ПРОАНАЛІЗОВАНИХ МЕТОДІВ РОЗПОДІЛУ НАВАНТАЖЕННЯ В МЕРЕЖІ.....	65
РОЗДІЛ 3 РОЗРОБКА МОДИФІКОВАНОГО МЕТОДУ РАЦІОНАЛЬНОГО РОЗПОДІЛУ НАВАНТАЖЕННЯ.....	67
3.1 Первинний збір інформації.....	67
3.2 Архітектура кластерного веб-серверу	68
3.3 Модифікований метод розподілу навантаження.....	70
3.4 Експериментальне дослідження ефективності запропонованого методу розподілу навантаження	73
РОЗДІЛ 4 РОЗРОБКА СТАРТАП-ПРОЕКТУ ЗА РЕЗУЛЬТАТАМИ ДИСЕРТАЦІЙНОЇ РОБОТИ.....	78
4.1 Опис ідеї стартап-проекту	78
4.2 Технологічний аудит ідеї проекту	79
4.3 Аналіз ринкових можливостей запуску стартап-проекту	80
4.4 Розроблення ринкової стратегії проекту.....	86
4.5 Розроблення маркетингової програми стартап-проекту	87
ЗАГАЛЬНІ ВИСНОВКИ ПО РОБОТІ	91
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	92

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

ЕОМ	Електронно-обчислювальна машина
ІКМ	Інфокомунікаційна мережа
КК	Комутація каналів
КП	Комутація повідомлень
ГК	Гібридна комутація
ТГ	Телеграфні канали
ТФ	Телефонний зв'язок
СП	Система передачі
АТМ	Asynchronous Transfer Mode
ЦП	Центральний процесор
МПД	Мережа передачі даних
FR	Frame relay
TCP	Transmission Control Protocol
MPLS	Multi Protocol Label Switching
PDH	Plesiochronous Digital Hierarchy
SDH	Synchronous Digital Hierarchy
ISO/OSI	International Standards Organization
MAC	Media Access Control
IP	Internet Protocol
LAN	Local Area Network
DNS	Domain Name System
NLB	Network Load Balancing
HTTP	HyperText Transfer Protocol
SMTP	Simple Mail Transfer Protocol
FIFO	First In, First Out
TLS	Transport Layer Security
RR	Round Robin

ВСТУП

Поширення різних обчислювальних пристроїв і розвиток мереж передачі даних, що зв'язують ці пристрої, змінило наш світ. Глибока інтеграція цієї розвиненої інфраструктури відіграє велику роль в повсякденному житті кожної людини. У світі, в якому розвиток інформаційних технологій набирає все більше обертів, важливу роль відіграють дані, а ще точніше процес обміну цих даних між різними компонентами мережі. Навіть для того, щоб знайти необхідну інформацію в мережі інтернет використовується велика кількість обчислювальних засобів. Одним із способів організації високопродуктивних обчислень є створення кластеру.

Кластер - це об'єднання декількох обчислювальних ресурсів в єдину систему. Це дозволяє отримати великий обчислювальний потенціал. У мережі кластер, це два або більше обчислювальних пристроїв, що працюють разом для загальної обчислювальної цілі. Окрім підвищеної обчислювальної потужності, загальні обчислювальні ресурси в кластері мережі також можуть забезпечити масштабованість, високу доступність, відмовостійкі можливості. Згруповані в локальну мережу декілька комп'ютерів можна назвати апаратним кластером, проте, сенс даного об'єднання - підвищення стабільності і працездатності системи за рахунок єдиного програмного забезпечення. Загалом, головним завданням кластера серверів – це уникнення «простою» системи. Тобто, будь-яке зовнішнє або внутрішнє втручання в мережу, яке може вивести із ладу хоча б один компонент мережі, повинен залишатися непоміченим для користувача.

В даний час у розробників кластерів є широкі можливості для вибору мережових технологій. Є багато прикладів побудови досить ефективних кластерів з використанням недорогого мережевого обладнання, яке можна побачити у звичайних локальних обчислювальних мережах. Але все ж таки важливу роль в побудові мережі та її налаштуванні відіграє саме розподіл навантаження в цій мережі.

В мережі інтернет стан перевантаження характеризується різким зростанням розмірів черг в окремих її вузлах. У разі, якщо механізм керування навантаження в мережі не може впоратись зі своїми завданнями, або ж він взагалі відсутній, може виникнути проблема в неможливості доступу користувачів до ресурсів окремих ділянок в мережі, а то й навіть - до виведення з ладу всієї мережі в цілому. В такому випадку мережа не може гарантувати необхідну якість обслуговування для вже встановлених з'єднань, а також і для встановлюючих з'єднань. Як правило, перевантаження може виникнути по різних причинах, наприклад, флуктуаціями потоків трафіку або виходом з ладу будь-якого мережевого компоненту, що може привести як до недотримання зобов'язань мережі щодо забезпечення якості обслуговування існуючих з'єднань, так і неможливістю встановлення нового з'єднання з запитаним якістю обслуговування.

Тому в даній роботі запропонований метод раціонального розподілу навантаження мережі, який дозволить вирішити проблеми, описані вище. Використовуючи запропонований метод, користувач матиме можливість налаштувати свою мережу, враховуючи свої потреби та вимоги до її роботи, а також раціонально розподіляти навантаження в мережі для підвищення якості обслуговування та її відмовостійкої і ефективної роботи.

РОЗДІЛ 1

ЗАГАЛЬНИЙ ОГЛЯД РОБОТИ ІНФОКОМУНІКАЦІЙНОЇ МЕРЕЖІ

У розділі проведений огляд структури інфокомунікаційної мережі, виділені основні її технологічні компоненти, а також розглянуті принципи побудови та роботи інфокомунікаційної мережі.

1.1 Загальні відомості та можливості мереж

Людство рухається по шляху створення глобального інформаційного суспільства. Його основою стає всесвітня інформаційна інфраструктура[1], складовою якої є потужні транспортні мережі зв'язку і розподілені мережі доступу, що надають інформацію користувачам. Інформаційна революція грає ключову роль у визначенні шляхів розвитку цивілізації. Це не просто технологічна, це соціо-технологічна революція, в ході якої формується новий домінуючий тип соціальної організації - інформаційне суспільство. Поточне тисячоліття увійде в історію людства як століття завершення переходу з індустріального суспільства в інформаційне. Процес переходу, який отримав назву інформатизації, інтенсивно реалізується у всіх країнах світу. В такому суспільстві будуть повністю задоволені інформаційні потреби населення.

Інформація - це дані, перетворені у більш корисну або зрозумілу форму. Іншими словами це сукупність даних, яка була організована для безпосереднього використання людства, оскільки інформація допомагає людям у процесі прийняття рішень. Інформація отримується шляхом складання елементів даних у значущу форму та може надалі оброблятися для формування знань. Сучасна цивілізація стала настільки складною і досконалою, що для виживання потрібно бути конкурентоспроможним. Це змушує людей інформувати себе про всі види подій у суспільстві.

Інформаційне суспільство - це таке суспільство, в якому виробництво і споживання інформації є найважливішим видом діяльності, а інформація визнається найбільш значущим ресурсом, нові інформаційні та телекомунікаційні технології і техніка стають базовими технологіями і технікою,

а інформаційне середовище разом з соціальної та екологічної - новими сферами проживання людини.

Інформаційне суспільство має такі відмінні ознаки

1. інформаційна економіка;
2. високий рівень інформаційних потреб усіх членів суспільства і фактичне їх задоволення для основної маси населення;
3. висока інформаційна культура;
4. вільний доступ кожного члена суспільства до інформації, обмежений тільки інформаційною безпекою особистості, суспільних груп і всього суспільства.

Інформаційному суспільству притаманне:

1. єдиний інформаційний простір;
2. домінування в економіці нових технологічних укладів, що базуються на масовому використанні мережових інформаційних технологій, перспективних засобів обчислювальної техніки і телекомунікацій;
3. провідна роль інформаційних ресурсів у забезпеченні сталого поступального розвитку суспільства;
4. зростання ролі телекомунікаційної інфраструктури в системі виробництва і посилення тенденцій до спільного функціонування в економіці інформаційних і грошових потоків;
5. фактичне задоволення потреб суспільства в інформаційних продуктах і послуг;
6. високий рівень освіти, обумовлений розширенням можливостей систем інформаційного обміну на міжнародному, національному та регіональному рівнях і, відповідно, підвищена роль кваліфікації, професіоналізму і здібностей до творчості як найважливіших характеристик праці;
7. висока значимість проблем забезпечення інформаційної безпеки особистості, суспільства і держави, наявність ефективної системи забезпечення

прав громадян і соціальних інститутів на вільне отримання, розповсюдження і використання інформації.

Інформаційне суспільство тісно зв'язане з провадженням інноваційних технологій, а також потребує безперервне інформування всередині інформаційного простору.

Інформаційний простір – це простір, в якому циркулюють інформаційні потоки даних. Розмір і топологічні властивості такого простору задаються інформаційною інфраструктурою.

Інформаційний потік - це інформація, яка переміщується в просторі і часі, а інформаційна інфраструктура – частина структури інформаційного простору, яке забезпечує створення і циркуляцію інформаційних потоків в просторі.

Інформаційна інфраструктура характеризується:

1. Якісним і кількісним складом елементів інфраструктури;
2. Інформаційної продуктивністю і пропускнуою спроможністю елементів всієї інформаційної інфраструктури в цілому.

До основних елементів інформаційної інфраструктури відносяться:

1. Телекомунікації;
2. Інформаційні ресурси;
3. Системи інформаційного обслуговування;
4. Системи забезпечення, розвитку і функціонування інформаційної інфраструктури.

Інформаційний ресурс - це інформація, придатна для задоволення інформаційних потреб та доступна будь-якій особі.

Інформаційний продукт - добре структурована, об'єктивна і досить повна інформація.

До інформаційних продуктів відносяться:

1. Інформація (дані, знання);
2. Носії інформації;
3. Інформаційні засоби техніки
4. Продукти, що забезпечують інформаційну діяльність.

Інформаційне суспільство в теоретичному плані характеризується високорозвиненою інформаційною сферою, в якій найважливішою діяльністю людини є створення, зберігання, обробка, обмін та накопичення інформації.

Інфокомунікація - це комплекс, який поєднує в собі сучасні інформаційні, комп'ютерні та телекомунікаційні технології, що реалізують системи і засоби, призначені для надання організаціям і населенню інформаційних та комунікаційних продуктів і послуг.

Інфокомунікаційна послуга - це послуга, що забезпечує задоволення телекомунікаційних або інформаційних, або тих і інших одночасно потреб споживача з наданням йому можливості керування процесом реалізації послуги.

Інформаційні процеси - збір, обробка, накопичення, зберігання, пошук і поширення інформації.

Інфокомунікаційні технології - це сукупність методів і способів обробки, накопичення, зберігання, відображення і забезпечення цілісності інформації, а також способів реалізації режимів її перенесення в просторі, що забезпечують певний гарантований рівень якості обслуговування.

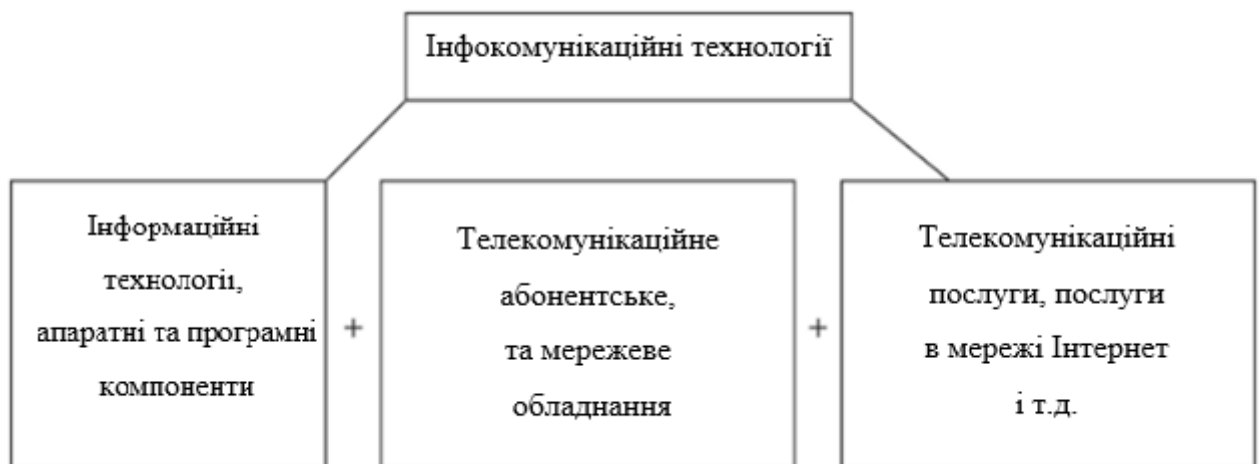


Рис. 1.1 Склад інфокомунікаційних технологій

Таким чином, згідно проведеного аналізу можна зробити висновок що інфокомунікаційна мережа - це технологічна(інфокомунікаційна) система, призначена для передачі по лініях зв'язку інформації, доступ до якої здійснюється з використанням засобів обчислювальної техніки.

Інфокомунікаційна система[2] - це сукупність інформаційної та телекомунікаційної систем. Інформаційна система включає в себе інформацію і користувача. Телекомунікаційна система забезпечує обмін інформації між джерелом та споживачем. Таким чином, інформаційно-комунікаційна система складається з сукупності телекомунікаційних мереж (телекомунікаційної підсистеми), прикладної підсистеми (засобів зберігання і обробки інформації, прикладних процесів та технологій), а також підсистеми джерел і споживачів інформації (призначені для користувача підсистеми).

Ядром інфокомунікаційної мережі є телекомунікаційні мережі(рис.1.2), до складу яких входять всі існуючі сьогодні та перспективні мережі електрозв'язку, іншою складовою інфокомунікаційної мережі є інформаційно-обчислювальні центри, бази даних та інші засоби зберігання, пошуку та обробки інформації. Третя складова - абонентські кінцеві пристрої (термінали), що забезпечують безпосередню взаємодію мережі з людиною.

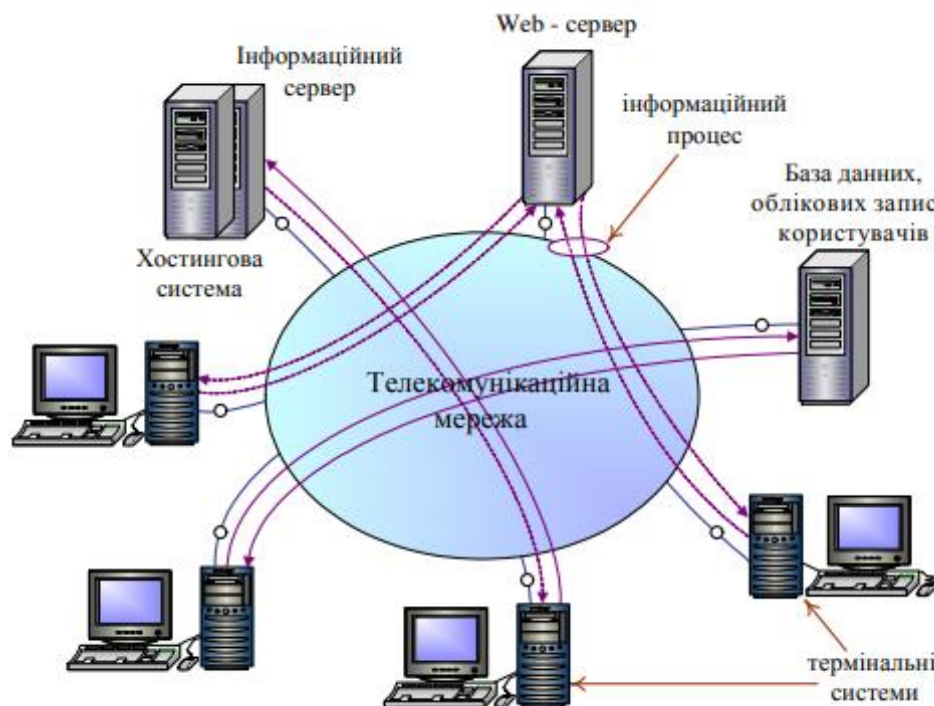


Рис. 1.2 Телекомунікаційна мережа

Традиційні мережі зв'язку (телефонні, телеграфні) відрізняються від інформаційних тим, що в першому випадку генерацією, прийомом, обробкою і зберіганням інформації займається людський мозок, а в інформаційних мережах частина цих функцій виконують обчислювальні машини. Характерно, що передача технічних засобів різних функцій по обробці інформації та навіть частини інтелектуальних функцій мозку є головним стимулом розвитку обчислювальної техніки і техніки зв'язку, в силу чого традиційні мережі зв'язку стають все більш схожими на комплекси ЕОМ, з'єднані каналами зв'язку. Тому і традиційні мережі зв'язку з їх інформаційними службами можна вважати як різновид інфокомунікаційних мереж. Всі існуючі і перспективні інфокомунікаційні мережі розрізняються, в першу чергу, різним рівнем інтеграції телекомунікаційних служб, починаючи від мереж для передачі одного виду інформації (телефонні, телеграфні, телевізійні та ін.) до широкосмугових мереж інтегрального обслуговування для передачі всіх видів інформації. Виходячи з цього інфокомунікаційні мережі можна класифікувати:

1. ІКМ для передачі одного або двох видів інформації (ІКМ - 0). До цього класу належать, в першу чергу, телеграфні мережі, які особливі тим, що мають автоматичну комутацію повідомлень, а також деякі телефонні мережі і мережі передачі даних, які обслуговують тільки один-два виду трафіку;
2. ІКМ з інтеграцією голосових, факсимільних служб і передачі даних на основі методу комутації каналів (ІКМ - 1). До цього класу належать існуючі аналогові і цифрові телефонні мережі з комутацією каналів та різних засобів інформаційних служб, які до них підключаються;
3. ІКМ з інтеграцією факсимільних, телеграфних, телеметричних служб і передачі даних на основі методу комутації пакетів (ІКМ - 2). До цього класу належать численні аналогові і цифрові мережі передачі даних X.25, FR і інші з підключеними до них засобами різних інформаційних служб.
4. ІКМ з інтеграцією голосових, факсимільних, телеграфних, телеметричних служб і передачі даних на основі методу комутації

пакетів(ІКМ – 3). До цього класу належать деякі мережі FR, АТМ та інші, а також Інтернет з реалізацією технології ІР-телефонії з інформаційними службами, які до них підключаються.

5. ІКМ-4, в яких разом із послугами ІКМ-3 забезпечується надання послуг з передачею відео та різного роду контенту (послуги Triple Play - одночасне постачання двох широкосмугових послуг (доступу до високошвидкісного інтернету і цифрового телебачення) і однієї вузькосмугової (телефону), через одну лінію широкосмугового зв'язку). Для їх побудови використовується ідеологія NGN, ядром яких є мультисервісні мережі високої пропускної здатності.
6. ІКМ-5 - подальший розвиток ІКМ-4, але основним видом трафіку в таких мережах стає телевізійна та інша відео-інформація. Пропускна здатність транспортних мереж ІКМ-5 повинна становити десятки і сотні Тбіт / с.

Таблиця 1.1

Порівняння класифікованих ІКМ по характеристикам

	Характеристики мереж	ІКМ-0	ІКМ-1	ІКМ-2	ІКМ-3	ІКМ-4	ІКМ-5
1	Метод комутації	КК, КП	КК, ГК	КП	КП	КП	КП
2	Тип трафіку	ТГ, ТФ дані	Аудіо, дані	Дані	Аудіо, дані	Відео, аудіо, дані	Відео, аудіо, дані
3	Пропускна здатність транспортних мереж	від 50 до 1200 біт/с	64 кбіт/с	до 2 Мбіт/с	до Гбіт/с	до десятків Гбіт/с	десятки Тбіт/с
4	Пропускна здатність абонентського доступу	від 50 до 1200 біт/с	64 кбіт/с	до сотень кбіт/с	до 2 Мбіт/с	до 2 Мбіт/с	до 30 Мбіт/с
5	Технології передачі і комутації	КК, КП	КК	X.25, FR, TCP/IP та ін.	FR, ATM, TCP/IP та ін.	ATM, TCP/IP, MPLS и та ін.	ATM, MPLS, нові
6	Транспортні технології		PDH, SDH	PDH, SDH	ATM, TCP/IP та ін.	ATM, MPLS та ін.	Нові оптичні

Отже, згідно проведеного аналізу можна зробити висновок, що інфокомунікаційна мережа являє собою сукупність кінцевих систем і будь-яких термінальних пристроїв користувачів, а також ресурсів мережі, які спільно

забезпечують виробництво і надання повного спектру телекомунікаційних та інформаційних послуг, які відповідають вимогам користувачів до їх якості. Інфокомунікаційні мережі призначені для забезпечення функціонування інформаційних систем і характеризуються великою різноманітністю комунікаційних послуг.

1.2 Структура мережі

Основне завдання мереж полягає в перенесенні інформаційних потоків між різними об'єктами (вузлами мережі). Мережі доступу забезпечують індивідуальну доставку інформаційних потоків на термінальне обладнання конкретних користувачів, тобто передачу інформації між інтерфейсом вузла, що надає послуги, і мережевим інтерфейсом кожного користувача. Сукупність транспортних мереж і мереж доступу утворюють національні, регіональні або місцеві інфокомунікаційні мережі.

Інфокомунікаційну мережу[8] можна розглядати як інформаційну систему, призначену для зберігання, передачі або обробки даних, накладену на мережу електрозв'язку, яка і здійснює передачу даних. У загальному вигляді сучасна телекомунікаційна мережа (мережа електрозв'язку) є ядром будь-якої інформаційної мережі, що забезпечує передачу і деякі види обробки даних.

Загалом на базі однієї телекомунікаційної мережі створюється цілий набір інформаційних (налагоджених) мереж різного призначення, що надають різноманітні послуги. Наприклад, мережа Інтернет існує поверх як телефонних, так і мереж передачі даних. Основна відмінність налагоджених мереж від мереж, побудованих на основі виділених каналів зв'язку, полягає в тому, що користувач отримує в своє розпорядження якийсь ресурс мережі з певними характеристиками (наприклад, гарантованою швидкістю передачі даних). Ключовим елементом інформаційно-комунікаційних мереж як і мереж електрозв'язку є термінальне обладнання.

Таблиця 1.2

Структура інформаційної системи

Термін	Визначення
--------	------------

Термінал	Пристрій введення / виведення даних і команд в інфокомунікаційних мережу, що взаємодіє безпосередньо з користувачем і перетворює дані в вид, придатний для передачі по мережах електрозв'язку.
Інтерфейс	Сукупність правил та засобів для взаємодії технологічних компонентів в мережі
Кінцевий пункт (endpoint)	Пункт, в якому розміщено термінальне обладнання користувачів і кінцеві системи мережі (сервери, на яких зосереджено інформаційні ресурси й застосування, у тому числі застосування системи керування мережею).
Канал передачі (transmission circuit)	Комплекс технічних засобів і середовища поширення для передачі сигналу в мережі електрозв'язку між мережними станціями, мережевими вузлами або між мережевою станцією і мережевим вузлом, а також між мережевою станцією або мережевим вузлом і кінцевим пристроєм первинної мережі
Лінія передачі (transmission line)	Сукупність лінійних трактів систем передачі і типових фізичних ланцюгів, що мають спільні лінійні споруди та обслуговують їх пристрої і одне і те ж середовище поширення в межах дії цих пристроїв
Абонентська лінія передачі (subscriber line)	Лінія передачі первинної мережі, що з'єднує мережеву станцію або мережевий вузол і кінцевий пристрій первинної мережі
Сполучна лінія (СЛ)	Лінія передачі, що з'єднує мережевий вузол і мережеву станцію. СЛ отримують назви в залежності від первинної мережі, до якої вона належить: магістральна, внутрішньозона, місцева
Вузловий пункт (node point), або просто вузол (node) мережі	Пункт мережі, в якому сходиться дві і більше СЛ і який є проміжним на шляху проходження потоків даних.

Приклад ієрархічної структури телекомунікаційної мережі, яка відображає інтенсивність трафіку між окремими її вузлами, розташованими в різних будівлях, населених пунктах і регіонах зображена на рисунку 3.

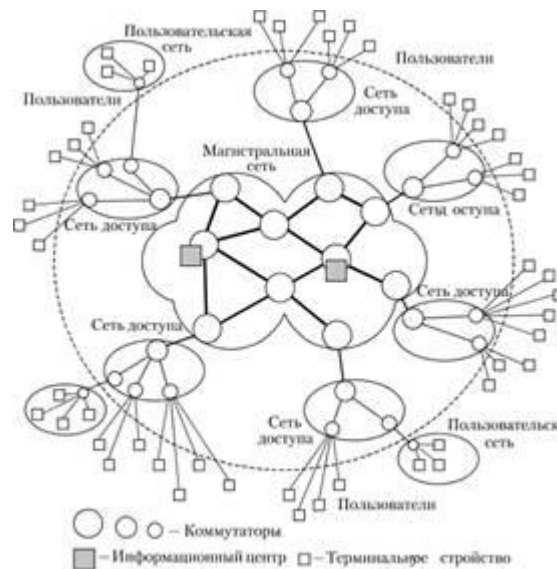


Рис. 1.3 Структура телекомунікаційної мережі

З малюнку структури телекомунікаційної мережі можна виділити основні структурні елементи мережі[4]:

1. Термінальні пристрої користувачів розташовані на периферії телекомунікаційної мережі і складають найнижчий рівень в її ієрархії. Зазвичай тип таких пристроїв визначає назву мережі. Основними термінальними пристроями в комп'ютерній мережі є комп'ютери, в телефонній - телефонні апарати, в телевізійній - телевізійні приймачі, в радіомовної мережі - радіоприймачі. Інформація від користувачів по абонентським каналах, часто званим абонентськими закінченнями, надходить на комутатори мережі.

2. Мережа доступу представляє наступний рівень ієрархії телекомунікаційної мережі. Велика така мережа може складатися з декількох рівнів. Основні функції мережі доступу складаються: в об'єднанні або мультиплексуванні інформаційних потоків, що надходять від численних користувальницьких пристроїв, в один загальний потік і передачі агрегованого потоку в комутатор магістральної мережі; виконує демультиплексування агрегованого потоку на окремі потоки таким чином, щоб на вхідний порт обладнання користувача надходила тільки адресована йому інформація.

3. Магістральна мережа призначена для транзиту агрегованих інформаційних потоків з мережі доступу від відправників в мережу доступу до

одержувачів. Вона містить комутатори і високошвидкісні лінії зв'язку (магістралі).

4. Інформаційний центр, або центр управління сервісами, призначений для надання інформаційних послуг користувачам (абонентам) мережі. Наприклад, інформаційні послуги Інтернету, а також телефонних мереж (отримання довідкової інформації, виклик швидкої допомоги і міліції) і мереж стільникового зв'язку (проведення телеголосування).

1.3 Основні технологічні компоненти мережі

Інфокомунікаційні мережі складаються, в основному, з семи типів структурних елементів:

1. Повторювачі (repeaters) працюють на фізичному рівні моделі ISO / OSI і зазвичай застосовуються в локальних мережах. Працюють як підсилювачі прийнятих сигналів і використовуються також для збільшення довжини сегмента в мережах топології «шина». На рис. 1.4 зображена типова схема: завдяки повторювачу довжина мережевий шини збільшується за рахунок сегмента 2;

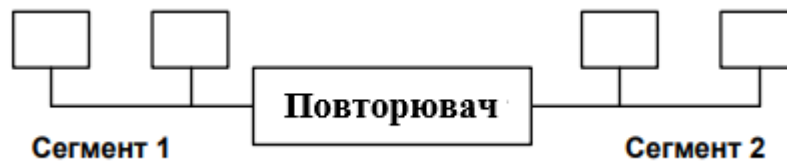


Рис. 1.4 Повторювач та сегменти

2. Мости (bridges) працюють на каналному рівні, що має підрівень MAC. У кожний вузол мережі входить плата мережевого інтерфейсу, що має унікальний MAC-адрес. Мости виділяють MAC-адреси з прийнятих кадрів даних і вибірково пересилають ці кадри в відповідні порти (рис.1.5). Міст ігнорує кадри, що передаються між вузлами, розташованими по одну сторону від нього. Наприклад, кадри, відправлені від вузла 1 до вузлів 2 і 3, міст пересилати не буде. Він відправляє кадри тільки вузлів, що знаходяться по інший бік від нього, наприклад, від вузла 1 до вузла 4 або 5. Одна з функцій моста - виділяти з кадрів вказівки по їх маршрутизації. Міст працює аналогічно маршрутизатора, різниця

тільки в тому, що міст працює на каналному рівні, а маршрутизатор - на мережевому. Міст здатний працювати як накопичувально-передавальний пристрій - приймати кадр цілком перед відправкою його в потрібний порт. При цьому кадр перевіряється за допомогою циклічної контрольної суми. Пошкоджені кадри не пересилаються. Це знижує навантаження на мережу, тому що пошкоджений кадр проходить тільки один сегмент і не потрапляє в інші;

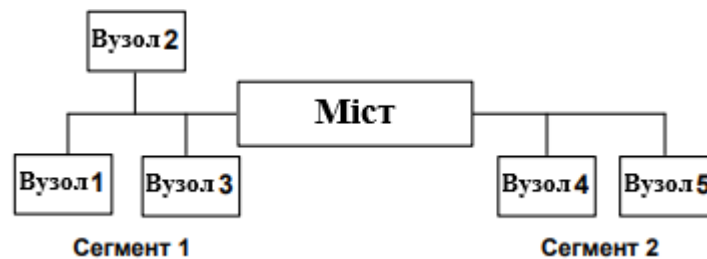


Рис. 1.5 Приклад роботи мостів

3. Маршрутизатор (routers) працюють на мережному рівні. На відміну від мостів, які пересилають пакети на основі таблиць фізичних адрес (наприклад, адрес Ethernet), маршрутизатори використовують таблиці логічних адрес (наприклад, IP-адрес). Вживається термін мультипротокольний маршрутизатор (multiprotocolrouter), який означає, що маршрутизатор розуміє кілька протоколів мережевого рівня, наприклад TCP і IPX. Розробники вітчизняної апаратури для пристроїв, аналогічних маршрутизаторів, використовували термін «центр комутації пакетів».

4. Комутатори (switch). Комутатор працює на каналному рівні майже як міст, але на відміну від моста він не накопичує кадр, а відразу після декодування адреси призначення відправляє його у відповідний порт. Цим досягається висока швидкість роботи комутатора. Недолік у тому, що комутатор пересилає всі кадри, навіть пошкоджені. Сучасні комутатори визначають адресу мережевого рівня, який різними способами зіставляється з портом комутатора і наступні пакети даних від того ж відправника до того ж одержувачу комутуються вже на каналному рівні, в той час як маршрутизатори роблять це на мережевому

рівні. Іншими словами, сучасний комутатор більше схожий на швидкісний маршрутизатор, а класичний комутатор - швидкісний міст.

5. Шлюзи (gateway) - це, як правило, апаратне і програмне забезпечення, що з'єднує дві різні мережі, в яких використовуються різні протоколи. Працюють на мережевому і більш високих рівнях. Так звані прикладні шлюзи при пересиланні даних з однієї мережі в іншу виконують трансляцію протоколів, як це робить поштовий шлюз, що конвертує два різних протоколу електронної пошти. Шлюз характеризується наявністю декількох адрес мережевого рівня, наприклад, кількох IP-адрес.

6. Хост (host) - це комп'ютер, на якому працює мережевий протокол, наприклад, TCP / IP. Хост обмінюється даними з іншими хост-комп'ютерами та значна частка діяльності в Інтернет обумовлена управлінням інформаційними потоками між хост-комп'ютерами. Типові приклади хостів: маршрутизатори, ПК, сервери, проксісервери, шлюзи і т. Д.

7. Вузли. Це, як правило, частина мережі, іншими словами будь-який пристрій, поєднаний з іншими пристроями в мережі.

Отже, будь-яка комп'ютерна мережа – це складний комплекс взаємопов'язаних і узгоджено функціонуючих програмних та апаратних компонентів.

1.4 Принципи побудови та роботи інфокомунікаційної мережі

В будь-якій інфокомунікаційній мережі взаємодія мережевих пристроїв реалізується через базову еталонну модель взаємодії відкритих систем (ISO/OSI)[9]. Модель OSI складається з семи рівнів. Кожний рівень виконує деякі процеси, апаратні і програмні засоби (об'єкти рівня), що реалізують функції по обробці і передачі даних. Кожен рівень OSI-моделі обслуговує суміжний старший рівень та відповідає за окремі специфічні функції в комунікаціях і реалізується відповідними технічними і програмними засобами мережі. Кожен рівень несе в собі якусь кількість інформації. Частина цієї інформації є службовою для цього рівня, наприклад, адреса.

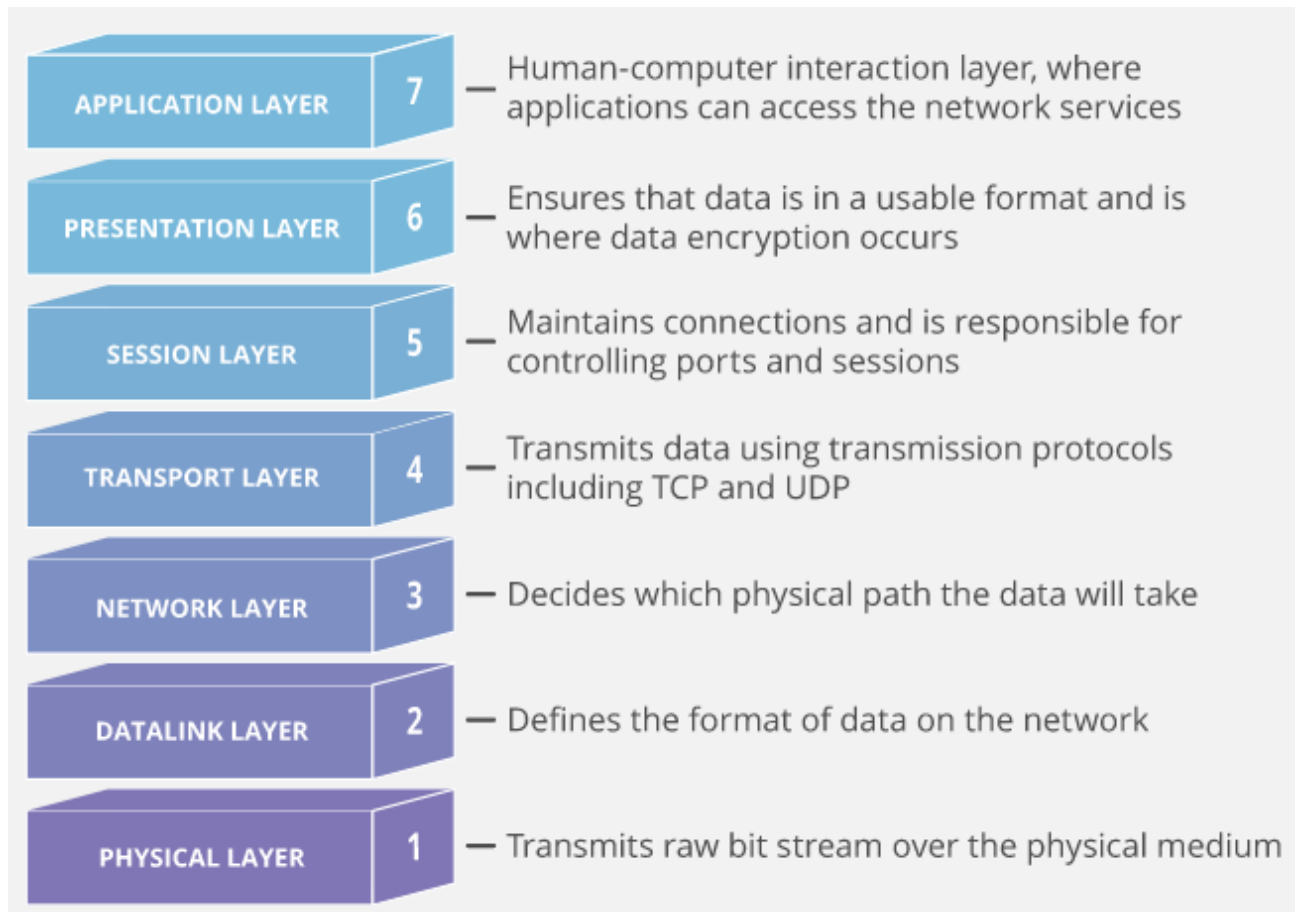


Рис. 1.6 Модель OSI

Модель OSI зображена на рисунку 1.6 має рівні:

1. Рівень 1 – Фізичний рівень(Physical layer). Цей рівень включає фізичне обладнання, яке бере участь у передачі даних, наприклад кабелі та комутатори. Це також рівень, на якому дані перетворюються в бітовий потік, який представляє собою рядок з одиниць та нулів. Фізичний рівень обох пристроїв також повинен узгоджувати узгодження сигналу, щоб можна було відрізнити одиниці від нулів на обох пристроях. Фізичний рівень реалізує управління каналом зв'язку - підключення і відключення каналу зв'язку, а також формування переданих сигналів. Отже, до фізичного рівня відносяться: синхронізація, кодування інформації, формування сигналів і передача біт по фізичних каналах зв'язку, модуляція. Фізичний рівень реалізується апаратно. При передачі даних на фізичному рівні не здійснюється контроль зайнятості

фізичної середовища передачі даних, перевірка доступності середовища передачі даних покладено на вищий рівень.

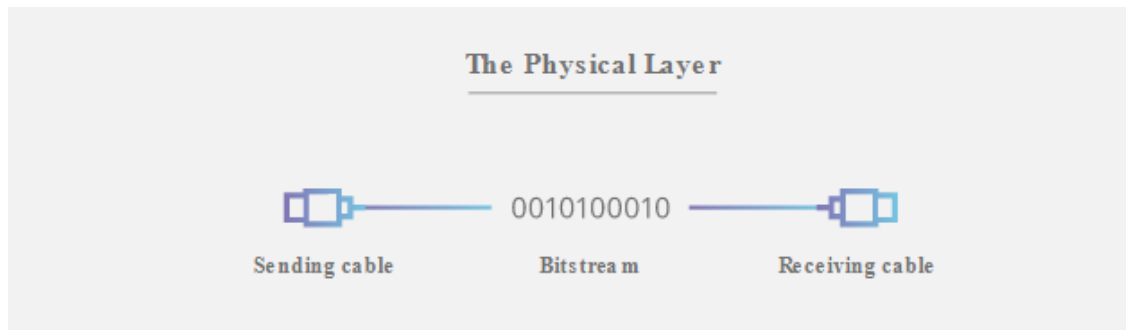


Рис. 1.7 Фізичний рівень моделі OSI

2. Рівень 2 - каналний рівень (Datalink Layer). Канальний рівень формує двосторонній канал зв'язку (тобто пряме налагодження зв'язку між суміжними вузлами мережі), використовуючи для цього два цифрових канали з протилежними напрямками передачі, які надаються першим рівнем (фізичним рівнем). Найважливіші функції рівня 2 - виявлення та виправлення помилок, які можуть виникнути на рівні 1, що робить незалежним якість послуг цього рівня від якості одержуваних «знизу» послуг передачі бітів. Канальний рівень надає наступні послуги або елементи послуг мережевого рівня - з'єднання каналного рівня, сервісні блоки даних каналного рівня, ідентифікатори кінцевого пункту з'єднання каналного рівня, також здійснює впорядкування блоків даних, здійснює оповіщення про помилки, керує потоком даних, визначає параметри якості послуг. На каналному рівні виконуються наступні функції - встановлення та розрив з'єднання каналного рівня, відображення сервісних блоків даних каналного рівня, розмежування і синхронізація, впорядкування блоків даних, виявлення помилок, відновлення при помилках, керування потоком даних, ідентифікація і обмін параметрами, керування перемиканням каналів даних, адміністративне управління каналним рівнем. Канальний рівень також призначений для виявлення помилок при передачі даних, біти, що передаються - групуються. Групи біт називаються кадрами (frames). Для ідентифікації кадру, початок і кінець кадру містить спеціальний набір бітів. У кадр включена

послідовність біт, яка називається контрольною сумою. Перед відправкою кадру обчислюється контрольна сума для даних, переданих в кадрі. Отримавши кадр, приймаюча сторона обчислює контрольну суму за тим самим алгоритмом, а потім порівнює її зі значенням контрольної суми, записаної в кадрі. Якщо значення контрольних сум збіглися - кадр переданий без спотворень, якщо ж ні - фіксується помилка для даного кадру. На каналному рівні можливе виправлення знайдених помилок за рахунок повторної передачі збійних кадрів. У деяких протоколах каналного рівня (наприклад, Ethernet) функція виправлення помилок відсутня, так як для каналного рівня функція виправлення помилок не є обов'язковою. Канальний рівень реалізується програмно-апаратно. Процедури каналного рівня - додавання в кадри відповідних адрес; Контроль помилок; Повторна, при необхідності, передача кадрів. Канальний рівень призначений також аби приховати від вищих рівнів подробиці технічної реалізації мережі. Таким чином, каналний рівень забезпечує створення, передачу і прийом кадрів даних, обслуговуючи запити мережевого рівня, використовуючи при цьому сервіс фізичного рівня. Функції каналного рівня реалізується мережевими адаптерами ПК і їх драйверами і різним комунікаційним обладнанням - мостами, маршрутизаторами, комутаторами.

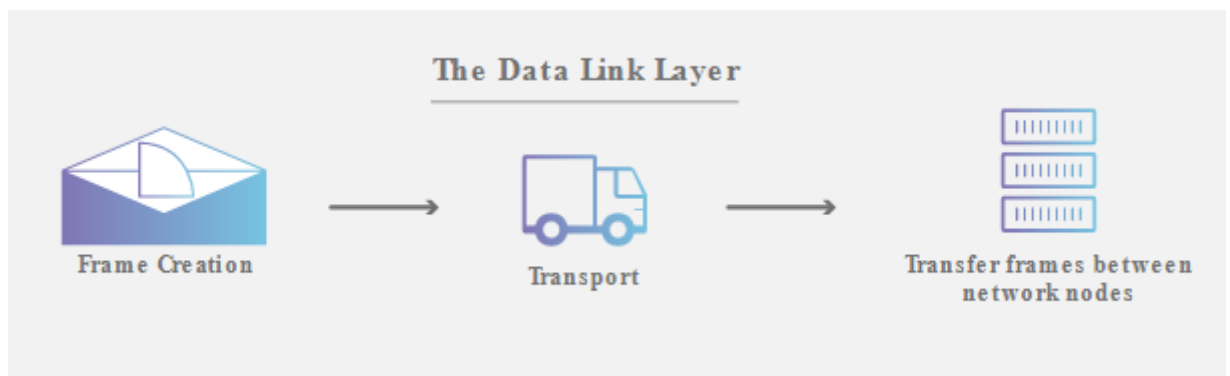


Рис. 1.8 Канальний рівень моделі OSI

3. Рівень 3 - Мережевий рівень. Цей рівень відповідає за передачу даних в мережі і формує мережеві послуги, маршрутизацію і комутацію з'єднань, що забезпечують передачу через всю мережу інформації, якою обмінюються

користувачі відкритих систем, розміщених в різних (і, в загальному випадку, несуміжних) вузлах мережі. Мережевий рівень форматує дані транспортного рівня і забезпечує їх мережевою адресою одержувача, необхідною для маршрутизації. Протоколи мережевого рівня призначені для утворення єдиної транспортної системи, що об'єднує декілька мереж з різними принципами передачі інформації між кінцевими вузлами, тобто всередині мережі отримання даних регулюється канальним рівнем, отримання даних між мережами - мережевим. Блок даних, який передається на мережевому рівні, називається пакетом (packet). Мережевий адреса - це специфічний ідентифікатор для кожної проміжної мережі між джерелом і приймачем інформації. Мережевий рівень реалізує - обробку помилок, мультиплексування пакетів, керування потоками даних. Мережевий рівень відповідає за адресацію (трансляцію фізичних і мережевих адрес, захисту мережевого взаємодії), пошук шляху від джерела до одержувача або між проміжними пристроями, встановлення і обслуговування логічного зв'язку між вузлами мережі. Таким чином, на мережевому рівні виконується трансляція логічних адрес та імен у фізичні адреси, а також визначаються найкоротші маршрути, виконується комутація та маршрутизація, відстеження неполадок і заторів в мережі.

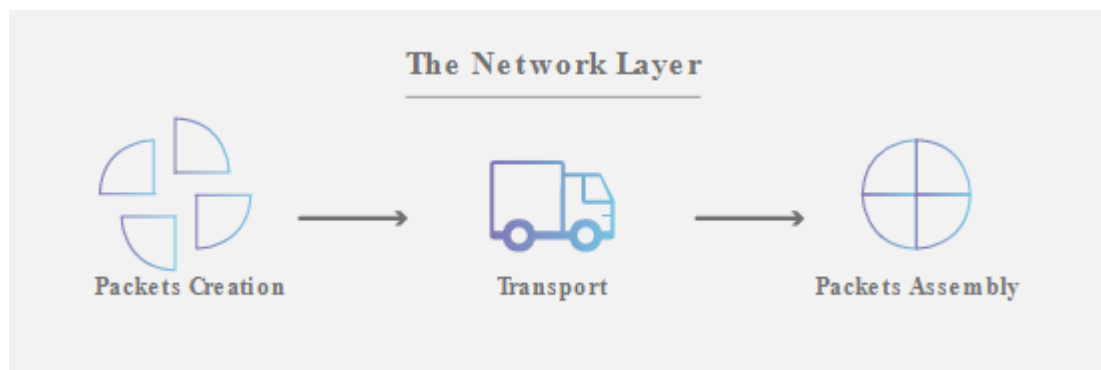


Рис. 1.9 Мережевий рівень моделі OSI

4. Рівень 4 - транспортний рівень (Transport Layer). Цей рівень керує наскрізною передачею повідомлень між кінцевими вузлами мережі, забезпечуючи надійність і економічну ефективність передачі даних незалежно

від користувача. При цьому кінцеві вузли мережі можуть взаємодіяти через кілька вузлів або навіть через кілька транзитних мереж. На транспортному рівні реалізується - перетворення довгих повідомлень в пакети при їх передачі в мережі і їх зворотнє перетворення, контроль послідовності проходження пакетів, регулювання трафіку в мережі, розпізнавання дублікатів пакетів і їх знищення. Транспортний рівень використовується додатком або верхнім рівнем стека прикладним і сеансовим для передачі даних з необхідним ступенем надійності. В OSI-моделі визначено п'ять класів сервісу, наданих транспортним рівнем. Для забезпечення певного рівня якості отримання інформації транспортний рівень виконує задачі - буферизація прийнятих пакетів, керуванням потоком, розподілом повідомлення сеансового рівня на пакети, їх нумерація, впорядковуванням пакетів, які надходять, адресацією прикладних процесів. Транспортний рівень може реалізовувати передачу даних без встановлення з'єднання або з попереднім встановленням з'єднання. В останньому випадку перед початком передачі даних з використанням спеціальних керуючих пакетів встановлюється з'єднання з транспортним рівнем процесу, якому призначені передані дані. Після того як всі дані передані, підключення закінчується. При передачі даних без встановлення з'єднання транспортний рівень використовується для передачі одиночних пакетів, дейтаграми, але не гарантуючи їх надійне отримання. Передача даних з встановленням з'єднання застосовується для надійної доставки даних.

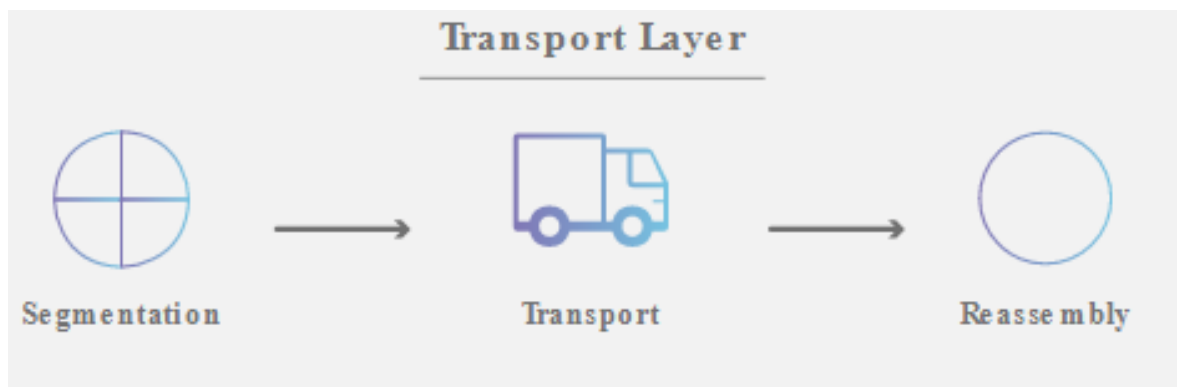


Рис. 1.10 Транспортний рівень моделі OSI

5. Рівень 5 – сеансовий рівень(The Session Layer). Це рівень, відповідальний за відкриття та закриття зв'язку між двома пристроями. Час між відкриттям та закриттям зв'язку називається сеансом. Рівень сеансу гарантує, що сеанс залишається відкритим достатньо довго для передачі всіх обмінюваних даних, а потім негайно закриває сеанс, щоб уникнути марної витрати ресурсів. Рівень сеансу також синхронізує передачу даних із контрольними точками. Наприклад, якщо передається 100-мегабайтний файл, сеансовий рівень може встановити контрольну точку кожні 5 мегабайт. У разі відключення або аварійного завершення роботи після передачі 52 мегабайт сеанс може бути відновлений з останньої контрольної точки, тобто потрібно передати лише ще 50 мегабайт даних. Без пунктів пропуску вся передача повинна була б початися спочатку. Сеансовий рівень рідко реалізується у вигляді окремих протоколів, функції цього рівня часто об'єднують з функціями прикладного рівня і реалізують в одному протоколі.

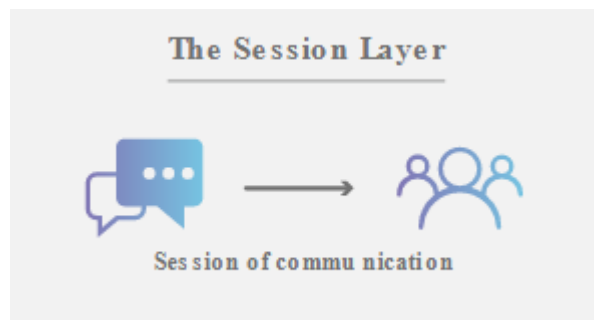


Рис. 1.11 Сеансовий рівень моделі OSI

6. Рівень 6 - рівень представлення. Цей рівень відповідає головним чином за підготовку даних, щоб вони могли бути використані рівнем програми; іншими словами, на цьому рівні обробляються дані в презентабельний вигляд для споживання додатками. Шар презентації відповідає за переклад, шифрування та стиснення даних. Два комунікаційні пристрої, що здійснюють зв'язок, можуть використовувати різні методи кодування, тому цей рівень відповідає за переклад вхідних даних у синтаксис, зрозумілий для приймаючого пристрою. Якщо пристрої здійснюють зв'язок за зашифрованим з'єднанням, рівень 6 відповідає

за додавання шифрування на кінці відправника, а також за декодування шифрування на кінці приймача, щоб він міг представити на рівні програми незашифровані, читабельні дані. Отже, рівень презентації також відповідає за стискання даних, які він отримує від рівня програми, перед тим, як доставити їх на рівень 5. Це допомагає поліпшити швидкість та ефективність зв'язку, мінімізуючи обсяг даних, які будуть передані.

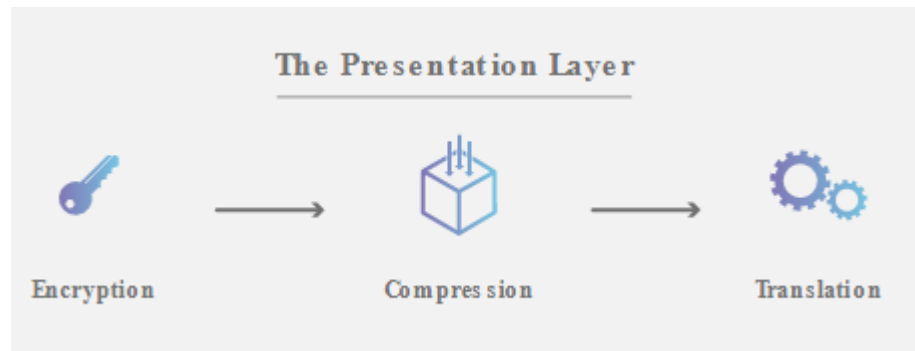


Рис. 1.12 Рівень представлення моделі OSI

7. Рівень 7 – Прикладний рівень. Це рівень, який безпосередньо взаємодіє з даними користувача. Такі програмні програми, як веб-браузери та поштові ящики клієнтів, покладаються на рівень додатків для ініціювання комунікацій. Але клієнтські програмні програми не є частиною прикладного рівня; швидше прикладний рівень відповідає за протоколи та маніпулювання даними, на які покладається програмне забезпечення, щоб представити значущі дані користувачеві. До протоколів прикладного рівня належать HTTP, а також SMTP (Simple Mail Transfer Protocol- один із протоколів, що забезпечує комунікацію електронною поштою). Прикладний рівень займається підтримкою прикладного процесу користувача та має справу з семантикою даних. Він є кордоном між процесами мережі і прикладними (для користувача) процесами. На цьому рівні виконуються обчислювальні, інформаційно-пошукові та довідкові роботи, здійснюється логічне перетворення даних користувача.

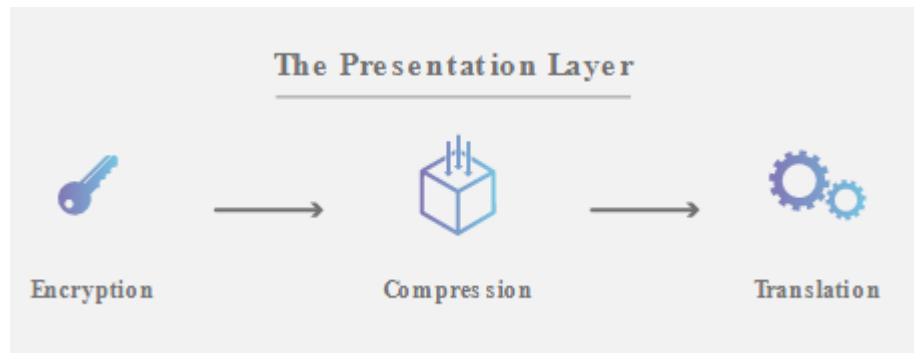


Рис. 1.13 Прикладний рівень моделі OSI

Таким чином, функції рівня 1-3 забезпечують транспортування інформації з одного пункту території в інший та пов'язані з окремими елементами мережевих зв'язків та з її внутрішньою структурою. Функції рівнів 4-7 відносяться лише до наскрізного зв'язку між кінцевими користувачами та визначаються таким чином, що вони не залежать від внутрішніх структур мережі. Залежно від набору технічних чи специфічних особливостей різних рівнів у них можуть бути сформовані та перероблені інформаційні блоки різних розмірів.

Висновки:

1. Проведений аналіз класифікації інфокомунікаційних мереж, оглянуті їх можливості та мережеві характеристики.
2. Розглянуто існуючі програмні та апаратні компоненти, які є невід'ємною частиною структури мережі, розглянута їх принципи роботи та можливості.
3. Розглянуті принципи побудови та роботи мережі на основі еталонної моделі OSI, проаналізована робота та значимість кожного її рівня.

РОЗДІЛ 2

ПЕРЕНАВАНТАЖЕННЯ МЕРЕЖІ ТА АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ РОЗПОДІЛУ НАВАНТАЖЕННЯ

У розділі виділені причини перенавантаження інфокомунікаційних мереж, а також виділені основні цілі розподілу навантаження в мережах. Проведений аналіз існуючих методів розподілу навантаження та сформований їх порівняльний аналіз.

2.1 Огляд та виділення основних причин перенавантаження в мережі

При проектуванні будь-яких інформаційних систем[19] однієї із суттєвих задач є планування ефективності. Адже будь-який інтернет-ресурс повинен забезпечувати гарну якість обслуговування, а також надавати можливість користувачам якомога швидше отримувати відповіді на запити до серверу. Це цілком легко досягти, застосовуючи більш потужне обчислювальне обладнання, як будь-який ресурс, воно має певну ціну. Тому в реальних умовах забезпечується знаходження розумного співвідношення між очікуваним навантаженням на проектуемую інформаційну систему та характеристик її апаратної частини. Дані збільшуються з кожною секундою в теперішній час, тому важливим питанням постає саме готовність будь-якого інтернет-ресурсу до перенавантаження серверу по різних причинах, а також при цьому можливості цього ресурсу, які не погіршували б якість обслуговування для користувачів.

Під перенавантаженням комп'ютерної мережі розуміється такий її стан, при якому мережеві ресурси на протязі достатньо тривалого інтервалу часу не здатні обробляти свої завдання. У комп'ютерних мережах на основі протоколу TCP / IP перенавантаження виникає через відсутність централізованого керування мережевими ресурсами. Наприклад, коли на маршрутизаторі або сервері поступає одночасно кілька TCP-пакетів, які повинні бути надіслані на одну адресу, або на один маршрутизатор (сервер) за короткий інтервал часу, за який надходить безліч пакетів від різних відправників. Після того, як пакети можуть бути оброблені лише послідовно, у разі одночасного введення декількох

пакетів, необхідне застосування спеціального механізму, який визначав би порядок обробки та передачі пакетів даних (пакети, які очікують обробки, можуть бути розміщені в спеціальній буферній пам'яті маршрутизатора (сервера) та передані для подальшої переробки відповідно до дисциплінарної FIFO). Застосування механізму буферної пам'яті в мережевих маршрутизаторах та серверах дозволяє обрати весь мережевий трафік при незначному збільшенні його інтенсивності. Однак, якщо інтенсивність передачі пакетів суттєво використовується, то враховуючи кінцевий розмір буферної пам'яті пакети будуть перевірятися.

Проблема перенавантаження не може бути розроблена реалізацією принципів необмеженою буферною пам'ятю, тому в цьому випадку також необмежено використовується довжина очікуваних пакетів і, зазвичай, необмежено збільшуються час передачі пакетів від відправників до отримувачів. А оскільки збільшується час обробки одного пакету даних, інші будуть в очікуванні і може виникнути процес «простою» всієї мережі, іншими словами мереже буде перенавантажена, в такому випадку пакет даних який не був відправлений отримувачу буде оброблятися мережевими компонентами спочатку. Таким чином, істотне збільшення розміру буферної пам'яті в маршрутизаторах і серверах є неефективним, так як у цьому випадку частина пакетів буде втрачена вже після того, як вони достатньо довгий час будуть оброблятися мережевими ресурсами.

Основною вимогою, якій має відповідати інформаційна мережа, є забезпечення користувачів ефективним доступом до ресурсів, які можуть розділятися (тобто колективного використовуватися). Усі інші вимоги – пропускна здатність, надійність, живучість – лише забезпечують якісне виконання цієї основної вимоги. Продуктивність інформаційної мережі як системи розподільчих ресурсів складається з оцінки таких параметрів, як час реакції мережі, затримка передачі і варіація затримки передачі даних, а також прозорість.

Час реакції мережі визначається як інтервал часу між поданням запиту користувача до певної мережевої служби (наприклад, передачі файлів) і отриманням відповіді на цей запит. Значення цього показника залежить від типу служби, до якої звертається користувач, від того, до якої категорії належить користувач та якою є продуктивність сервера, куди він звертається, а також від ступеня завантаженості елементів мережі, через які проходить його запит.

Затримка передачі визначається як час між моментом надходження пакету даних на вхід будь-якого мережевого пристрою або фрагмента мережі і моментом виходу з неї. Цей параметр по суті характеризує етапи тимчасової обробки пакетів при проходженні їх мережею. При цьому продуктивність мережі оцінюється, як правило, максимальною затримкою передачі та варіацією затримки передачі даних.

Варіація затримки характеризує коливання затримки в часі. Великий діапазон в значеннях затримки негативно позначається на якості наданої користувачеві інформації при передаванні чутливих до затримки видів трафіку, таких як відеодані, тобто мультимедійний трафік. Це супроводжується виникненням різних причин, наприклад, нерозбірливістю мови, поганою якістю зображення та ін.

Прозорість характеризується властивістю мережі приховувати від користувача принципи її внутрішньої організації. Користувач не повинен знати місцезнаходження програмних та інформаційних ресурсів. Для роботи з віддаленими ресурсами мережі він повинен використовувати ті ж самі команди й процедури, що й для роботи з ресурсами свого комп'ютера. Вимога до прозорості забезпечує користувачам зручність і простоту роботи в мережі.

Передача даних в комп'ютерній мережі передбачає організацію фізичного або логічного з'єднання між взаємодіючими користувачами мережі (кінцевими вузлами). Організація взаємодії між абонентами комп'ютерної мережі називається комутацією. Комутація в мережі може бути реалізована різними способами (рис.1.14), які можна розбити на дві групи:

1. Способи комутації без проміжного зберігання даних;

2. Способи комутації зі зберіганням даних в проміжних вузлах.

Як спосіб комутації без проміжного зберігання даних в комп'ютерних мережах застосовується комутація каналів, використовувана в традиційних телефонних мережах зв'язку. Для передачі даних в комп'ютерних мережах був розроблений новий спосіб комутації - комутація повідомлень, що припускає використання в якості вузлів зв'язку спеціалізованих засобів обчислювальної техніки, що дозволяло реалізувати в проміжних вузлах зберігання переданих даних, що забезпечує ряд переваг в порівнянні з комутацією каналів. Подальший розвиток способів комутацій був направлений на вдосконалення комутації повідомлень для забезпечення певної якості передачі даних.



Рис. 2.1 Способи комутації в мережах для передачі даних

2.1.1 Причини перенавантаження використовуючи комутацію каналів в мережі

Комутація каналів заснована на формуванні єдиного фізичного з'єднання (каналу) між взаємодіючими абонентами для безпосередньої передачі даних від краю до краю так само, реалізація в мережах (рис 2.2).

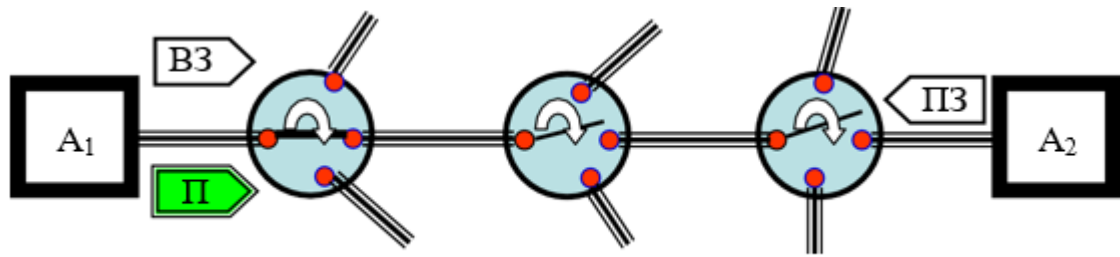


Рис. 2.2 Схема комутації каналів

Якщо абонент A_1 хоче передати дані абоненту A_2 , то перед початком передачі він попередньо повинен встановити з'єднання з абонентом A_2 шляхом використання спеціального службового повідомлення «ВЗ - встановити з'єднання», яке «прокладає» шлях, формуючи в кожному з проміжних вузлів безпосереднє фізичне (електричне) з'єднання між вхідним і вихідним портами вузла. Після того, як службове повідомлення досягне абонента A_2 , останній компонент формує і посилає по створеному шляху (маршруту) абоненту A_1 нове службове повідомлення «ПЗ - підтвердити з'єднання», яке підтверджує встановлення з'єднання між абонентами мережі. Тільки після отримання такого повідомлення абонент A_1 може почати передачу повідомлення «П» абоненту A_2 за встановленим маршрутом. Створене фізичне з'єднання зазвичай існує протягом часу передачі даних, званого сеансом або сесією (session), по завершенню якого ця сполука може бути зруйновано. Такий канал між двома абонентами мережі називається тимчасовим або комутованим, на відміну від некомутованого (виділеного) каналу, який формується один раз і існує постійно або, принаймні, протягом тривалого часу, незалежно від того, передаються дані або ж канал простоює. У найпростішому випадку вузол мережі з комутацією каналів можна розглядати як перемикач, що забезпечує в кожен момент часу електричне з'єднання між двома портами (точками входу і виходу) вузла. У телефонній мережі такими «перемикачами» є автоматичні телефонні станції (АТС).

До основних переваг комутації каналів належать:

1. Можливість використання існуючих і досить добре розвинених мереж зв'язку;

2. Висока ефективність при передачі великих обсягів даних, оскільки в цьому випадку відносне значення накладних витрат на встановлення з'єднання виявляється незначним;
3. Відсутність необхідності в зберіганні переданих даних в проміжних вузлах мережі.

При цьому комутація каналів має наступні недоліками:

1. Канали зв'язку на всьому шляху передачі даних повинні мати однакові пропускні спроможності і забезпечувати однакову швидкість передачі даних, в іншому випадку, якщо пропускна здатність деякого каналу зв'язку виявиться менше пропускної здатності попереднього каналу, відбудеться втрата переданих даних, оскільки в проміжних вузлах відсутня можливість буферизації (тимчасового зберігання) даних;
2. Великі накладні витрати на встановлення з'єднання на початковому етапі, що негативно позначається при передачі невеликих обсягів даних, оскільки в цьому випадку відносне значення накладних витрат на встановлення з'єднання виявляється істотним, що призводить до неефективного використання ресурсів (пропускної спроможності) каналів зв'язку, що проявляється в значному зменшенні реальної швидкості передачі даних по відношенню до максимально можливої швидкості каналу, званої пропускною спроможністю.
3. Телефонні канали зв'язку, орієнтовані на передачу голосу, мають порівняно низьку якість і забезпечують передачу комп'ютерних даних з невисокою швидкістю, що не дозволяє їх використовувати в високошвидкісних магістральних мережах.

Альтернативою комутації каналів, що усуває властиві їй недоліки, є комутація повідомлень.

2.1.2 Причини перенавантаження використовуючи комутацію повідомлень в мережі

Комутація повідомлень, на відміну від комутації каналів, передбачає зберігання переданих повідомлень в буферній пам'яті проміжних вузлів, що знаходяться на шляху передачі, який прокладається в кожному вузлі відповідно

до заданого алгоритму маршрутизації (рис.2.3). При цьому не потрібно попередньо встановлювати з'єднання між взаємодіючими абонентами.

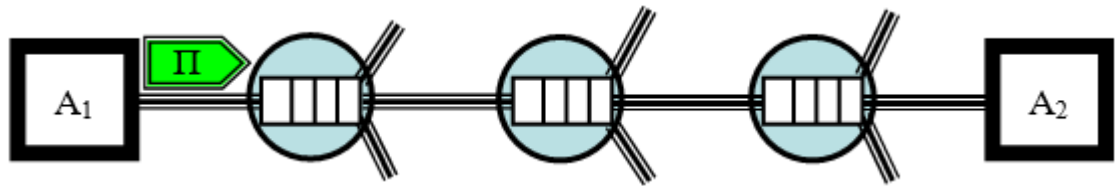


Рис. 2.3 Схема комутації повідомлень

Якщо абонент A_1 бажає передати повідомлення Π абоненту A_2 то він, не встановлюючи безпосереднє з'єднання з A_2 , посилає повідомлення до вузла зв'язку, до якого він підключений. Там повідомлення зберігається в буфері вузла протягом деякого часу, необхідного для аналізу заголовка, визначення відповідно до заданого алгоритму маршрутизації наступного вузла і, можливо, очікування звільнення каналу зв'язку з цим вузлом, якщо канал зайнятий передачею раніше обробленого повідомлення. Проходячи таким чином через всі вузли, що знаходяться на шляху передачі, повідомлення досягає кінцевого абонента A_2 . При цьому, напрямок передачі повідомлення, тобто його маршрут в мережі, визначається тільки після надходження повідомлення в той чи інший вузол мережі, а не встановлюється заздалегідь, як це відбувається при комутації каналів. Завдяки такій організації передачі даних між взаємодіючими абонентами, комутація повідомлень володіє наступними перевагами в порівнянні з комутацією каналів:

1. Не потрібно попереднє встановлення з'єднання, що істотно знижує накладні витрати, але не робить їх нульовими, оскільки є непродуктивні витрати часу в кожному вузлі на обробку заголовка і реалізацію алгоритму маршрутизації; проте в цілому ці витрати істотно менше в порівнянні з витратами на встановлення з'єднання при комутації каналів;

2. канали зв'язку на всьому шляху передачі можуть мати різні пропускні спроможності, оскільки буферизація повідомлень в вузлах мережі дозволяє налагодити різницю в пропускних можливостях вхідного і вихідного каналів вузла.

Недоліками комутації повідомлень є:

1. Необхідність зберігання переданих повідомлень в проміжних вузлах, що вимагає значної ємності буферної пам'яті, яка розраховується як добуток ємності одного буфера на максимально можливу кількість повідомлень, які одночасно можуть перебувати в вузлі; ємність одного буфера повинна бути розрахована на повідомлення максимальної довжини, яка, наприклад, для відео файлів може становити кілька гігабайт, що робить ємність буферної пам'яті вузла не виправдано великий; при цьому коефіцієнт використання (завантаження) буферної пам'яті виявляється незначним, оскільки більшість повідомлень, займаючи один буфер, матимуть довжину багато менше, ніж ємність буфера;

2. затримка в проміжних вузлах може виявитися чималою, особливо через великий час очікування звільнення вихідного каналу зв'язку при великому завантаженні мережі, що призводить до збільшення часу доставки повідомлень;

3. монополізація середовища передачі (каналу зв'язку) на тривалий проміжок часу при передачі довгих повідомлень призводить до не виправдано великих затримок коротких повідомлень в зв'язку з очікуванням звільнення каналу, тривалість якого може багаторазово перевищувати час безпосередньої передачі цих повідомлень.

2.1.3 Причини перенавантаження використовуючи комутацію пакетів в мережі

Комутація пакетів відрізняється від комутації повідомлень лише тим, що кожне повідомлення в мережі розбивається на блоки фіксованої довжини L_p const (Крім останнього блоку: $L_1 \leq L_p$, де L_1 -довжина при комутації повідомлень), званих пакетами (рис 2.4), кожен з яких має структуру аналогічну структуру повідомлень: заголовок, текст і, можливо, кінцеву точку. При цьому, заголовки всіх пакетів одного і того ж повідомлення містять одні й ті ж адреси призначення і джерела. Кожен пакет повідомлення передається в мережі як незалежний блок даних відповідно до адреси призначення, зазначених у заголовку.

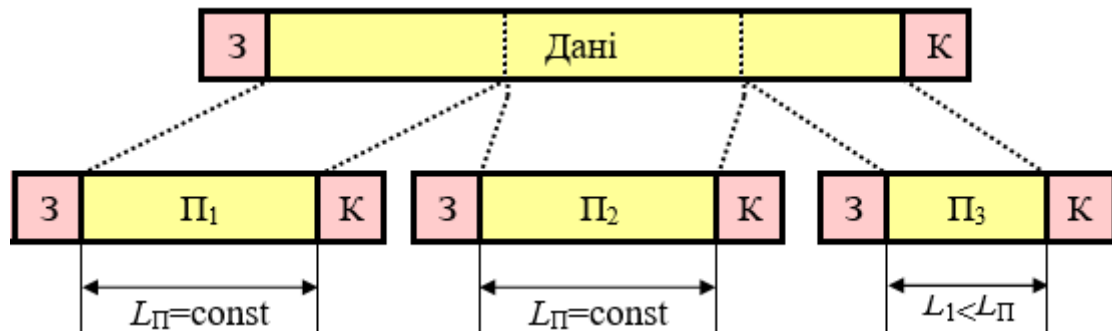


Рис. 2.4 Схема комутації пакетів

Комутація пакетів в порівнянні з комутацією повідомлень дозволяє реалізувати більш ефективну передачу даних за рахунок наступних властивих їй переваг:

1. менший час доставки повідомлення в мережі;
2. більш ефективне використання буферної пам'яті в вузлах;
3. більш ефективна організація надійної передачі даних;
4. середовище передачі не монополізується одним повідомленням на тривалий час;
5. затримка пакетів в вузлах менше, ніж затримка повідомлень.

Затримка пакетів в вузлах менше, ніж затримка повідомлень, яка складається з наступних складових:

1. Прийом (запис) надходить блоку даних (пакета або повідомлення) у вхідний буфер вузла;
2. Підрахунок і перевірка контрольної суми блоку даних
3. Передача блоку даних з вхідного буфера у вихідний буфер
4. Очікування звільнення вихідного каналу, зайнятого передачею раніше отриманих блоків даних
5. передача даних у вихідний канал зв'язку і звільнення вихідного буфера вузла

Незважаючи на очевидні переваги, комутації пакетів притаманні недоліки, які полягають у наступному:

1. великі накладні витрати на передачу і аналіз заголовків всіх пакетів повідомлення, що знижує ефективну (реальну) пропускну здатність каналу зв'язку, яка використовується безпосередньо для передачі даних, і, отже, збільшує час доставки повідомлення в мережі, в тому числі і за рахунок додаткових витрат часу на обробку заголовків пакетів у вузлах мережі.

2. необхідність складання з пакетів повідомлення у вузлі призначення може істотно збільшити час доставки повідомлення кінцевому абоненту за рахунок очікування приходу всіх пакетів повідомлення, оскільки в разі втрати хоча б одного пакету, повідомлення не зможе бути зібрано в кінцевому вузлі мережі; при цьому виникає серйозна проблема, пов'язана з визначенням гранично допустимого часу очікування пакетів для збирання повідомлення в кінцевому вузлі; при великому значенні цього часу в кінцевому вузлі може зібратися велика кількість пакетів різних повідомлень, що призведе до переповнення буферної пам'яті вузла і, як наслідок, до втрати переданих пакетів або до відмови в прийомі нових пакетів, що, в свою чергу, не дозволить зібрати повідомлення, маленьке значення гранично допустимого часу очікування пакетів для збирання повідомлення в кінцевому вузлі може створювати таку ситуацію, при якій велика кількість повідомлень не зможуть дочекатися приходу останнього пакета і, оскільки після закінчення цього часу всі пакети таких повідомлень будуть видалені з буферної пам'яті, буде потрібно повторна передача всіх пакетів цих повідомлень, що призведе до значної завантаженні обладнання (вузлів і каналів) мережі і, в межі, може викликати перевантаження мережі.

2.1.4 Причини перенавантаження використовуючи комутацію комірок в мережі

Комутація комірок - спосіб комутації, який можна розглядати як окремий випадок комутації пакетів зі строго фіксованою довжиною переданих блоків даних в 53 байта (рис 2.5).

3	Данные
5	48 байт

Рис. 2.5 Приклад виконання комутації комірок

Перші комп'ютерні мережі будувалися для передачі цифрових (Комп'ютерних) даних з однією вимогою - забезпечити надійністю (без помилок) доставку даних, при цьому час доставки не був критичним. Розвиток комп'ютерних технологій і поява необхідності передачі мультимедійних даних, таких як мова та відео, виокремив, поряд з надійною доставкою, нову вимогу до передачі даних в комп'ютерних мережах: мінімізація часу доставки повідомлень. Для реалізації цієї концепції була розроблена мережева технологія, що отримала назву Asynchronous Transfer Mode (ATM) - режим асинхронної передачі, призначення якої - передача мультимедійних даних в комп'ютерній мережі з мінімальною затримкою. При розгляді комутації пакетів, чим коротше пакети, тим менше час доставки всього повідомлення. Виходячи з цього, в АТМ-мережах в якості одиниці передачі даних був обраний блок розміром в 53 байта (5 байт - заголовок і 48 байт - дані), названий коміркою даних. Такий дивний розмір комірки даних з'явився в результаті компромісу двох протидіючих груп, з яких одна група (за однією з версій: традиційні зв'язківці - телефоністи) наполягала на меншому значенні поля даних в 32 байта, а інша (комп'ютерники) - на значенні в 64 байта. Менше значення розміру комірок забезпечило б менші затримки при доставці даних, однак при цьому зростають накладні витрати на передачу заголовків комірок даних, що знижує корисну (ефективну) пропускну здатність середовища передачі. Беручи до уваги, що в мультимедійних мережах зазвичай використовуються високошвидкісні канали, втрати пропускну здатності можуть виявитися значними, що негативно позначиться на економічній ефективності комп'ютерної мережі.

Підводячи підсумок сказаному, можна відзначити такі переваги комутації осередків:

1. маленькі затримки комірок даних (не монополізується канал зв'язку);
2. швидка обробка заголовку комірки даних в вузлах, оскільки місце розташування заголовка суворо фіксоване;
3. більш ефективна, у порівнянні з комутацією пакетів, організація буферної пам'яті і надійної передачі даних;

Основним недоліком комутації комірок даних є - наявність порівняно великих накладних витрат на передачу заголовку (майже 10%) і, як наслідок, значна втрата пропускної здатності, особливо в разі високошвидкісних каналів зв'язку.

Отже, комутація пакетів і комутація каналів - основні способи передачі даних в комп'ютерних мережах, оскільки комутація пакетів забезпечує більш ефективну передачу даних через МПД в порівнянні з комутацією повідомлень (в першу чергу, значно менші затримки), а комутація каналів може бути досить легко реалізована на основі існуючої інфокомунікаційної мережі.

Згідно проведеного аналізу перенавантаження в мережі можна виділити основні причини виникнення перенавантаження в мережі:

1. Суттєве збільшення обсягу передачі даних по одному мережевому каналу (через одночасну роботу кількох користувачів);
2. ТСП-протокол повторно передає пакети, які вже знаходяться в процесі передачі або навіть прийнятих користувачем (що можливо, якщо час передачі перевищує деяке критичне значення), при такому збільшенні фізичної пропускної здатності каналів зв'язку є лише тимчасовим рішенням цієї проблеми
3. втрачені пак*ети, які деякого часу передавались через мережу, але не були передані отримувачам, тобто мережеві ресурси були зайняті даними пакетами фактично даремно.
4. фрагментація пакетів, коли в мережі передаються фрагменти, які будуть в подальшому видалені, так як з них неможливо відновити вихідний пакет

Результатом перегляду в комп'ютерній мережі є високий рівень втрати пакетів, суттєве збільшення часу їх доставки від відправників до одержувачів і

навіть тимчасової недоступності мережевих ресурсів та сервісів. Перенавантаження в каналі зв'язку при виконанні процедур аутентифікації може сприяти неможливості реалізації даних засобів захисту інформації.

Таким чином, цілісною актуальною є розробка та застосування засобів управління мережевим трафіком, які дозволяють уникати потенційних перевантажень у каналах зв'язку.

2.2 Основні цілі розподілу навантаження в інфокомунікаційних мережах

У термінології комп'ютерних систем збалансування навантажень або вирахування навантажень (англ. Load balancing) - це метод розподілу завдань між кількома мережевими пристроями (наприклад, серверами) з метою оптимізації використання ресурсів, скорочення часу обслуговування запитів, горизонтального масштабування кластера (динамічне додавання / видалення пристроїв), а також забезпечення відмовостійкості (резервування). У комп'ютерних балансувачах навантаження розподіляється між кількома обчислювальними ресурсами, такими як комп'ютери, комп'ютерні кластери, мережі, центральні процесори або диски. Ціль розподілу навантаження[3] - оптимізація використання ресурсів, максимізація пропускної здатності, зменшення часу відгуку та запобігання перенавантаження якого-небудь одного ресурсу. Використання кількох компонентів для розподілу навантаження замість одного компонента може підвищити надійність і доступність за рахунок резервування. Розподіл навантаження передбачає зазвичай наявність спеціального програмного забезпечення або апаратних засобів, таких як багаторівневий комутатор або система доменних імен, як серверний процес.

Розподіл навантаження відрізняється від фізичного з'єднання тим, що збалансованість навантажень ділить трафік між мережевими інтерфейсами на мережевому рівні (моделі OSI рівень 4) на основі, в той час як з'єднання каналів передбачає розподіл трафіку між фізичними інтерфейсами на більш низькому рівні, або в пакеті (модель OSI рівень 3) або по каналу зв'язку (модель OSI рівень 2). Необхідність керування трафіком в мережі обумовлена такими

особливостями, притаманними мережевого трафіку сучасних комп'ютерних мереж:

1. неоднорідність трафіку, що характеризується наявністю в мережі декількох типів даних, які можна розділити на дві великі групи: мультимедійні (мова, аудіо та відео) і комп'ютерні (електронні листи, файли і т.д.);
2. наявність різних (диференційованих) вимог до якості передачі даних різних типів;
3. випадковий характер і нестаціонарність мережевого трафіку, обумовлені зміною інтенсивностей потоків даних в різний час доби і непередбачуваністю характеру і темпу роботи користувачів у комп'ютерній мережі;
4. в свою чергу, нестаціонарність мережевого трафіку може призвести до виникнення в комп'ютерній мережі періодів перевантажень і навіть до блокувань.

Блокування в мережі можуть виникнути в результаті заповнення буферної пам'яті вузлів. Найпростіший приклад блокувань показаний на рис. 2.6, де буфери двох сусідніх вузлів, потребуючих обмінятися пакетами, заповнені до кінця.



Рис. 2.6 Приклад блокування в мережі через переповнення буферної пам'яті вузлів

Це призводить до ситуації, коли обмін пакетами неможливий, незважаючи на те, що в принципі буферної пам'яті досить для зберігання наявних пакетів. Однак, для того щоб прийняти пакет від сусіднього вузла, необхідно мати хоча б один вільний буфер. Таким чином, вузли виявляються заблокованими, що може, в кінцевому рахунку, привести до зупинки (блокування) всієї мережі. На рис.2.7 показано залежність продуктивності мережі передачі даних, яка вимірюється кількістю пакетів, переданих в мережі за одиницю часу, від кількості пакетів, що знаходяться в мережі.

Спочатку продуктивність мережі передачі даних, як і слід було очікувати, зростає із збільшенням кількості знаходяться в мережі пакетів M , досягаючи при $M_{\text{пред}}$ до певного граничного значення, що представляє собою пропускну здатність мережі. При цьому завантаження, принаймні, одного з вузлів або каналів зв'язку, званого вузьким місцем, досягає при M , $M_{\text{пред}}$ значення, що призводить до перевантаження мережі. Подальше збільшення кількості пакетів в мережі не призводить до зростання продуктивності, значення якої буде визначатися продуктивністю вузького місця. Більш того, подальше збільшення кількості пакетів в мережі веде до зниження продуктивності і навіть до припинення передачі пакетів, тобто до зупинки мережі, що, зокрема, пов'язано з виникаючими в мережі блокуваннями.



Рис. 2.7 Залежність продуктивності мережі від кількості пакетів

Перераховані вище особливості комп'ютерних мереж обумовлюють необхідність управління неоднорідним трафіком в мережі для вирішення наступних завдань:

1. Забезпечення надійної передачі даних, яка передбачає доставку даних абоненту без втрат і без спотворення даних (за рахунок застосування механізмів квітування і тайм-ауту).
2. забезпечення ефективного завантаження дорогого мережевого устаткування (каналів і вузлів) мережі (за рахунок реалізації механізму перерозподілу потоків даних в процесі адаптивної маршрутизації)

3. малі затримки при передачі по мережі повідомлень і, перш за все, мультимедійних (за рахунок маршрутизації і пріоритетів)

4. запобігання перевантажень і блокувань при передачі даних (за рахунок пріоритетів і обмеження входить до мережі трафіку).

Виходячи з цього аналізу можна виділити основні цілі розподілу навантаження в інфокомунікаційних мережах:

1. Підвищення ступеня утилізації ресурсів
2. Підвищення пропускнуої здатності
3. Зменшення часу відгуку системи
4. Боротьба з відмовами і перевантаженнями

Отже, завдяки наявності в групі однотипних елементів, в неї можна додавати нові елементи або виключати наявні, щоб привести загальну продуктивність до бажаного значення. Використання декількох елементів замість одного підвищує відмовостійкість системи, так як вихід з ладу одного елемента не зупинить роботи інших і групи, в цілому. Присутність декількох елементів в групі дозволяє планово перерозподілити навантаження між елементами так, щоб з'явилася можливість виконати налагодження або обслуговування одного з них без припинення роботи всієї комп'ютерної системи.

2.3 Аналіз існуючих методів розподілу навантаження в ІКМ

2.3.1 Статичний розподіл навантаження

Статичний розподіл навантаження не бере до уваги стан системи для розподілу завдань. Таким чином, стан системи включає такі параметри, як рівень завантаження (а іноді навіть перевантаження) певних процесорів. Замість цього заздалегідь робляться припущення для всієї системи, такі як час прибуття і вимоги до ресурсів для вхідних завдань. Крім того, відомо кількість процесорів, їх потужність і швидкість передачі даних. Отже, статичне балансування навантаження спрямоване на те, щоб пов'язати відомий набір завдань з доступними процесорами та мінімізувати певну функцію продуктивності.

Методи статичного балансування навантаження зазвичай централізовані навколо маршрутизатора або моста, який розподіляє навантаження і оптимізує

функцію продуктивності. Ця мінімізація може враховувати інформацію, що відноситься до завдань, які повинні бути розподілені, і визначати очікуваний час виконання. Перевага статичних алгоритмів в тому, що їх легко налаштувати і вони надзвичайно ефективні в разі досить регулярних завдань (наприклад, обробка HTTP-запитів з веб-сайту). Проте, існує деяка статистична різниця в розподілі завдань, яка може призвести до перевантаження деяких обчислювальних пристроїв.

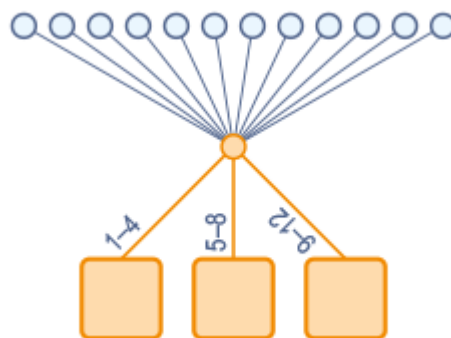


Рис. 2.8 Приклад виконання статичного розподілу навантаження

На прикладі виконання показано, що користувачів в мережі можна направляти з одного регіону на один сервер, а з іншого - на інший. При статичному розподілі прив'язка користувача до конкретного сервера постійна, що не змінюється ні за яких умов.

2.3.2 Динамічний розподіл навантаження

На відміну від алгоритмів статичного розподілу навантаження, динамічні алгоритми враховують поточне навантаження кожного з обчислювальних блоків (також званих вузлами) в системі. При такому підході завдання можна динамічно переміщати з перевантаженого вузла на недовантажений, щоб прискорити обробку. Хоча ці алгоритми набагато складніше розробити, вони можуть дати відмінні результати, зокрема, коли час виконання сильно варіюється від одного завдання до іншого. При динамічному балансуванні навантаження архітектура мережі може бути більш модульної, оскільки не обов'язково мати певний вузол, виділений для розподілу роботи. Коли завдання однозначно призначаються процесору відповідно до його станом в даний момент, це унікальне призначення.

Якщо, з іншого боку, завдання можуть бути постійно перерозподілені відповідно до стану системи і її розвитком, це називається динамічним призначенням. Очевидно, що алгоритм балансування навантаження, який вимагає занадто багато комунікацій для прийняття рішень, ризикує уповільнити рішення загальної проблеми. В якості іншого засобу розподілу навантаження використовують спеціалізовані програмні або апаратні балансувачі. Вони можуть забезпечити різні алгоритми розподілу запитів між серверами, в тому числі, такі, при яких враховується поточний стан серверів або інших важливих факторів. Наприклад, програмний балансувач HTTP-запитів часто використовує різного роду проксі-сервери, що перенаправляють запити за встановленими правилами куди слід(приклад виконання на рис. 2.9).

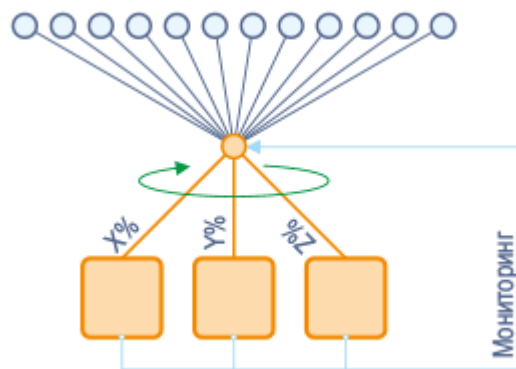


Рис. 2.9 Приклад виконання динамічного розподілу навантаження

2.3.3 Розподіл навантаження на транспортному рівні

TCP Load Balancer використовує протокол управління передачею (TCP), який перевіряє інформацію, що надсилається на адреси інтернет-протоколу (IP). Це забезпечує надходження даних без помилок до додатків, що не належать до HTTP.

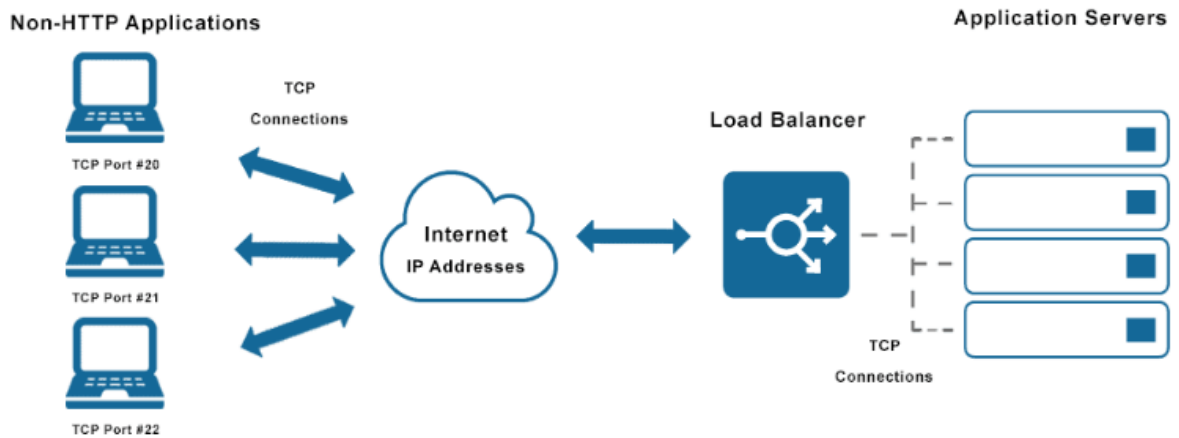


Рис. 2.10 Приклад виконання розподілу навантаження на транспортному рівні

TCP балансувач навантаження - це тип балансувача навантаження, який використовує протокол управління передачею (TCP), який працює на рівні 4 - транспортному рівні - в моделі взаємозв'язку відкритих систем (OSI). Трафік TCP спілкується на проміжному рівні між прикладною програмою та Інтернет-протоколом (IP). Конфігурація балансування навантаження TCP забезпечує надійний і перевірений помилками потік пакетів на IP-адреси, які в іншому випадку легко втратити або пошкодити. Кожній програмі присвоюється унікальний номер порту TCP, щоб забезпечити доставку до правильної програми та забезпечити перевірку працездатності. TCP балансувач навантаження призначений для програм, які не використовують HTTP. При розгортанні перед кластером баз даних балансування навантаження TCP розподіляє запити по всіх доступних конфігураціях сервера. Будь-які дані програми, призначені для сервера, пересилаються на доступний сервер через нове TCP-з'єднання. Поділ (або проксі-сервер) клієнта на підключення до сервера дозволяє підвищити рівень безпеки. Це також забезпечує кращу продуктивність сервера та підключення клієнта. В порівнянні з балансувачем TCP балансувач навантаження HTTP - це проста архітектура запитів / відповідей HTTP для трафіку HTTP. Але балансувач навантаження TCP призначений для програм, які не розмовляють HTTP.

Балансувач навантаження[21] програми може працювати на рівні TCP, поки він знаходиться на 4 рівні моделі взаємозв'язку відкритих систем (OSI). Це пояснюється тим, що балансування навантаження TCP призначені для програм, які не використовують HTTP. Наприклад, рівень 7 у моделі OSI очікує, що весь мережевий трафік буде HTTP. Але балансувач навантаження, що працює на рівні 4, зчитує інформацію про пакет TCP і направляє зв'язок у потрібне місце. Під “рівнем” балансування навантаження розуміється, наскільки далеко до мережевого стеку повинна пройти комунікація, перш ніж вона буде спрямоване до кінцевого пункту призначення. Маршрутизатор балансування навантаження програми, заснований на TCP, має меншу затримку, оскільки мережевий зв'язок не повинен проходити весь шлях вгору і назад по мережевому стеку на балансувачі навантаження. Цей вид балансування є найпростішим, оскільки клієнт звертається до балансувача, той перенаправляє запит одного з серверів, який і буде його обробляти. Вибір сервера, на якому буде оброблятися запит, може здійснюватися відповідно різних алгоритмів - шляхом простого кругового перебору(Round Robin), шляхом вибору найменш завантаженого сервера з пулу і т.п.

2.3.4 Розподіл навантаження на прикладному рівні

Прикладне балансування схоже на транспортне[30] – але при використанні проксі-сервера, і запити користувачів передаються на ці сервери. Але, на відміну від транспортного балансування, прикладна розподіляє навантаження з урахуванням запитуваних сторінок, контенту або дій. Наприклад, запис на сайт (створення облікових записів і подібні дії) будуть проводитися через один сервер, а читання - через інший. Можна розподілити навантаження по типу контенту (аудіо, відео, зображення, текст).

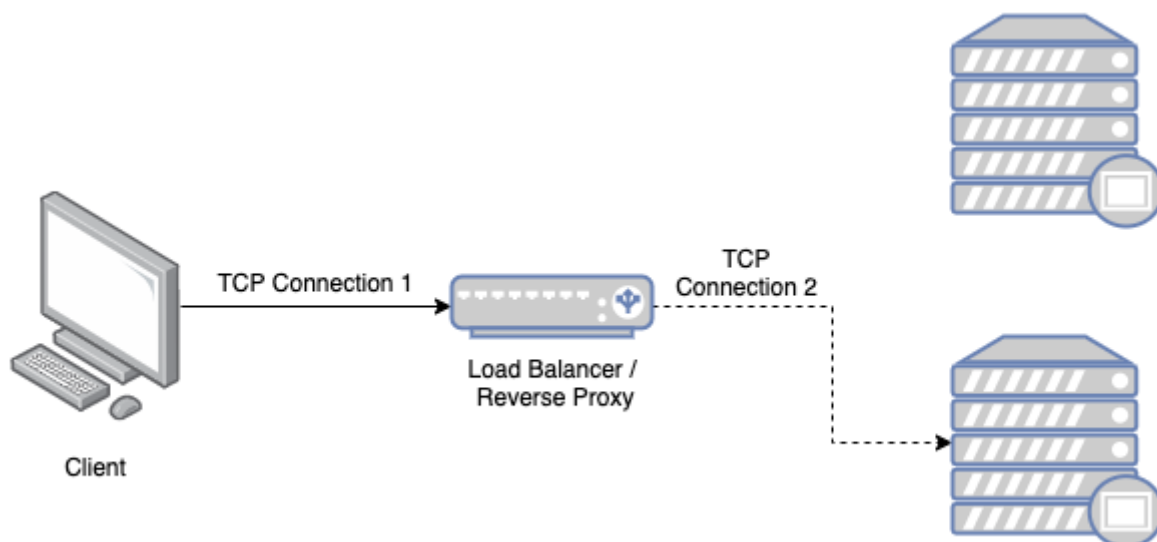


Рис. 2.11 Приклад виконання розподілу навантаження на прикладному рівні

Коли клієнт відправляє запит до серверу, робота прикладного рівня полягає в створенні TCP-з'єднання з балансувачем навантаження. Потім Load Balancer створює нове TCP-з'єднання з одним із вихідних серверів. Таким чином, є 2 TCP-з'єднання порівняно з 1 у TCP / UDP-прохідному L4(транспортний рівень моделі OSI) балансувач навантаження. На прикладному рівні, відомі дані, що містяться в запиті, це дозволяє виконувати різні операції, наприклад:

1. Аутентифікація - 401, якщо якийсь заголовок відсутній;
2. Розумна маршрутизація - маршрутизація / виклик платежів до певного вищого потоку;
3. Припинення TLS

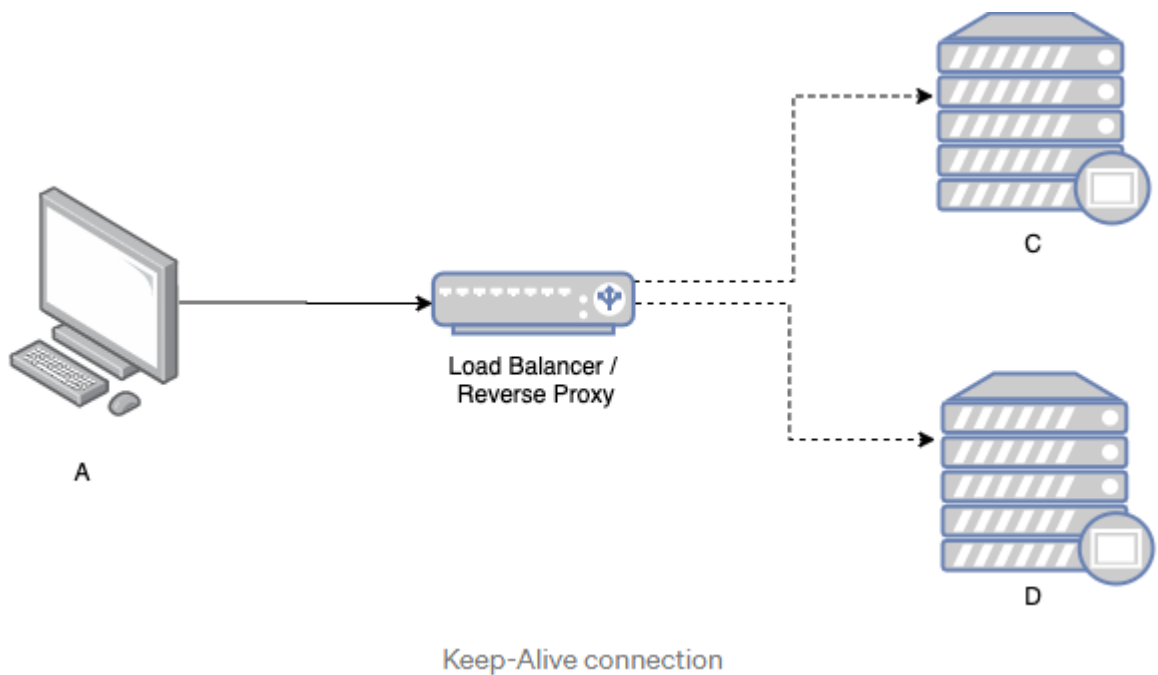


Рис. 2.12 Приклад виконання розподілу навантаження на прикладному рівні

У випадку мультиплексних протоколів / підтримання життя, балансувач навантаження на рівні 7(моделі OSI) створює TCP-з'єднання з кожним вихідним потоком для одного підключення клієнта, а не вибирає тільки один вихідний потік. Це означає, що коли А(рис 2.12) створює зв'язок з балансувачем навантаження, балансувач навантаження створює два з'єднання, одне з С і одне з D. Балансувач навантаження на рівні 7 є комбінованим балансувачем з іншими рівнями моделі OSI, а саме 5(сеансовий), 6(представлення).

2.3.5 Розподіл навантаження на мережевому рівні

Цей метод передбачає використання набору фізичних серверів[26]. Це досить дорогий, але дуже ефективний метод, який надійно захищає ІТ-спеціаліста від перевищення навантаження на сервер. Сенс методу полягає в тому, що кілька різних машин відповідають за роботу одної ІР-адреси. Балансування на мережевому рівні передбачає вирішення наступного завдання: потрібно зробити так, щоб за один конкретний ІР-адрес сервера відповідав різним фізичним машинам. Така балансування може здійснюватися за допомогою безлічі різноманітних способів:

1. DNS-балансування. На одне доменне ім'я виділяється кілька IP-адрес. Сервер, на який буде спрямований клієнтський запит, зазвичай визначається за допомогою алгоритму Round Robin;
2. Побудова NLB-кластера. При використанні цього способу сервери об'єднуються в кластер, що складається з вхідних і обчислювальних вузлів. Розподіл навантаження здійснюється за допомогою спеціального алгоритму.
3. Балансування по IP з використанням додаткового маршрутизатора
4. Балансування за територіальною ознакою здійснюється шляхом розміщення однакових сервісів з однаковими адресами в територіально різних регіонах. Балансування за територіальною ознакою також використовується в багатьох CDN.

2.3.6 Round Robin

Назва цього алгоритму походить від принципу кругового руху, коли кожна людина отримує рівну частку отримання відповіді на запит по черзі. Цей метод здебільшого використовується для багатозадачності. У плануванні методу Round Robin[27] кожне готове завдання виконується по черзі лише в циклічній черзі протягом обмеженого часу. Також завдяки цьому методу можна виконувати процеси без зупинок.

Характеристики Round Robin:

1. Процесор переміщується до наступного процесу після фіксованого інтервал часу, який називається часовим квантом.
2. Відхилений процес додається до кінця черги.
3. Часовий фрагмент повинен бути мінімальним, який призначається для конкретного завдання, яке потрібно обробити.
4. Це алгоритм реального часу, який реагує на події протягом певного періоду часу.

Алгоритм Round Robin – це удосконалення методу FCFS. Метод FCFS – це алгоритм планування використання ОС, яка автоматично виконує запити і процеси в черзі по мірі їх надходження. Приклад виконання роботи методу FCFS зображено на рис 2.13.

Process	Duration	Oder	Arrival Time
P1	24	1	0
P2	3	2	0
P3	4	3	0

Gantt Chart :



Рис. 2.13 Приклад виконання методу FCFS

З прикладу видно що процесу P1 – потрібно 24, наприклад, мілісекунди на виконання, відповідно процесу P2 - 3, P3 – 4. Згідно прикладу перший пріоритет має процес P1, відповідно P2 – другий, P3 – третій. Спочатку виконуватиметься процес P1, оскільки має перший пріоритет, процес P2 буде на черзі, оскільки має другий пріоритет. Процес P2 буде в очікуванні до того моменту, поки не процес P1 не будет виконаний, тобто процес P2 буде чекати свого виконання 24 мілісекунди і тільки після того почне виконуватись, аналогічно і з процесом P3.

Метод Round Robin заключається в тому що кожний із процесів виконується тільки за відповідний період часу(квант часу). Наприклад, є три процеси P1(перший в черзі) – якому потрібно 4 мілісекунди на виконання, P2(другий в черзі) – 3 мілісекунди, P3(третій в черзі) – 5 мілісекунд на виконання(рис 27), при цьому квант часу буде 2 мілісекунди.

Process Queue	Burst time
P1	4
P2	3
P3	5

Рис. 2.14 Умови прикладу виконання Round Robin

Кроки виконання методу Round Robin:

1. Виконання починається з процесу P1, який має перший пріоритет та час – 4 мілісекунди. Тут кожен процес виконується протягом 2 секунд(квант часу). P2 і P3 все ще в черзі очікування.

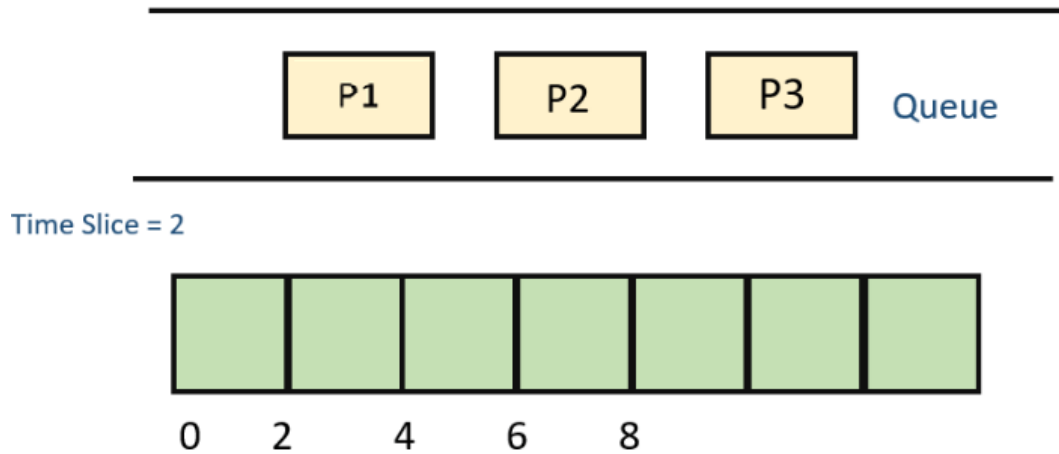


Рис. 2.15 Перший крок виконання Round Robin

2. Після проходження 2 мілісекунд, P1 додається до кінця черги, і P2 починає виконуватися.

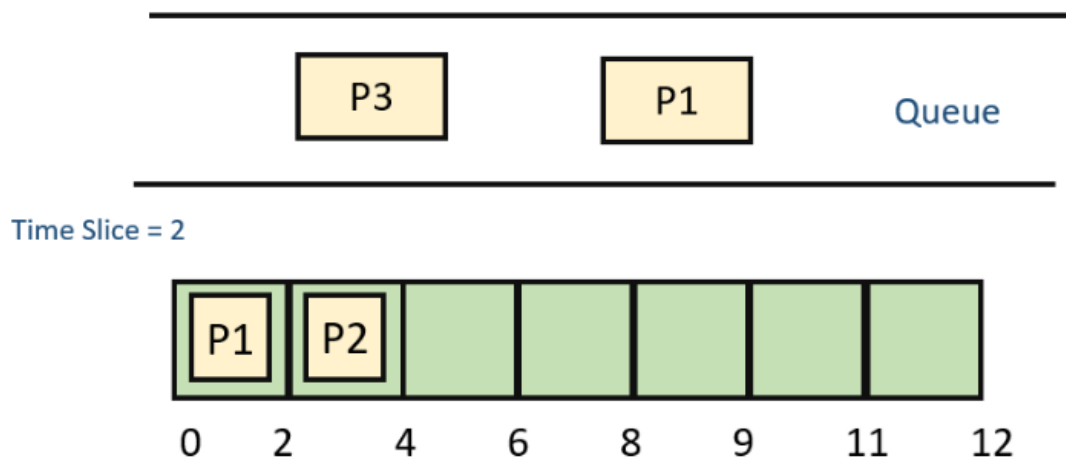


Рис. 2.16 Другий крок виконання Round Robin

3. В момент часу 4мілісекунди, P2 вимикається і додається в кінець черги. P3 починає виконуватися

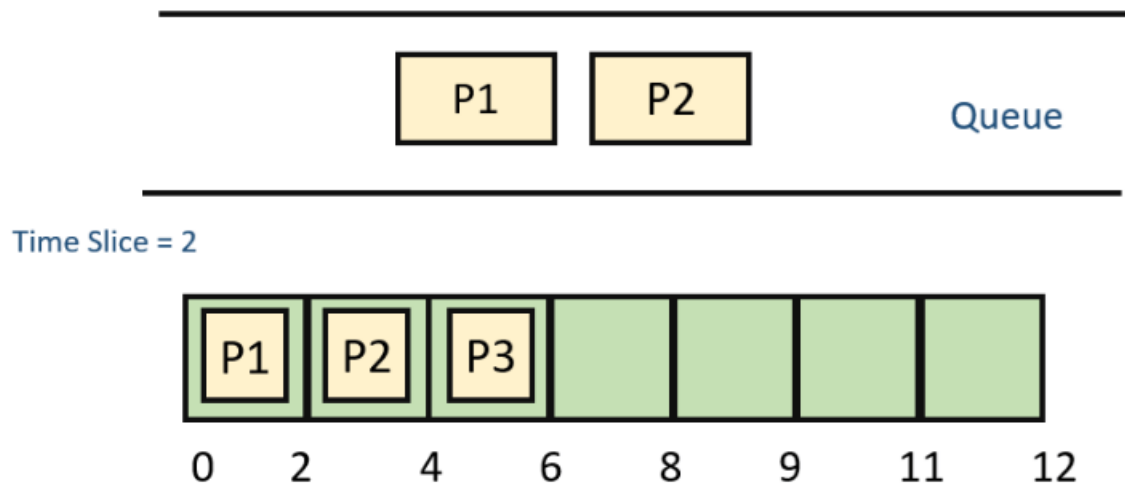


Рис. 2.17 Третій крок виконання Round Robin

4. В момент часу = 6, P3 виключається і додається в кінець черги. P1 починає виконуватися знову.

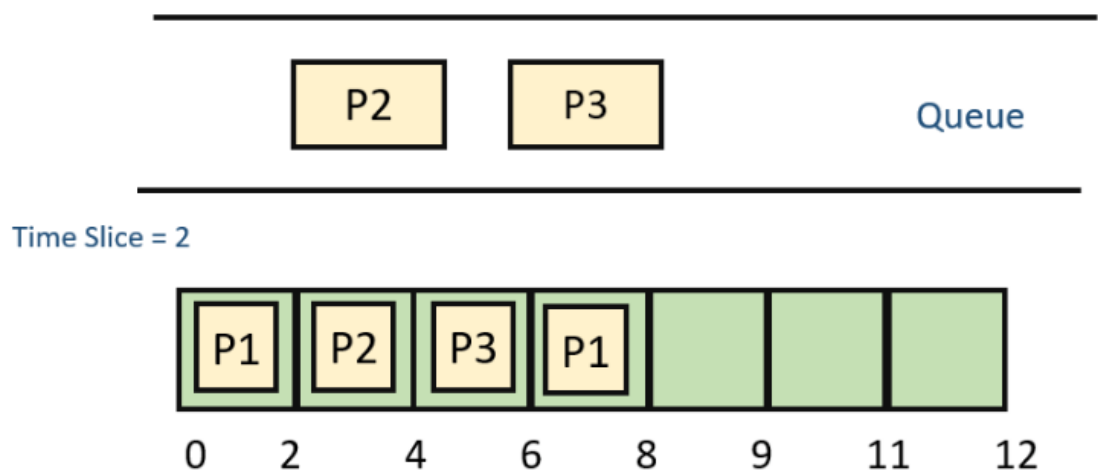


Рис. 2.18 Четвертий крок виконання Round Robin

5. У момент часу 8 мілісекунд, час P1 має час на виконання 4 мілісекунди, оскільки він пройшов вже два цикли (два рази виконувався з квантом часу – 2 мілісекунди), він завершив своє виконання. P2 починає виконання.

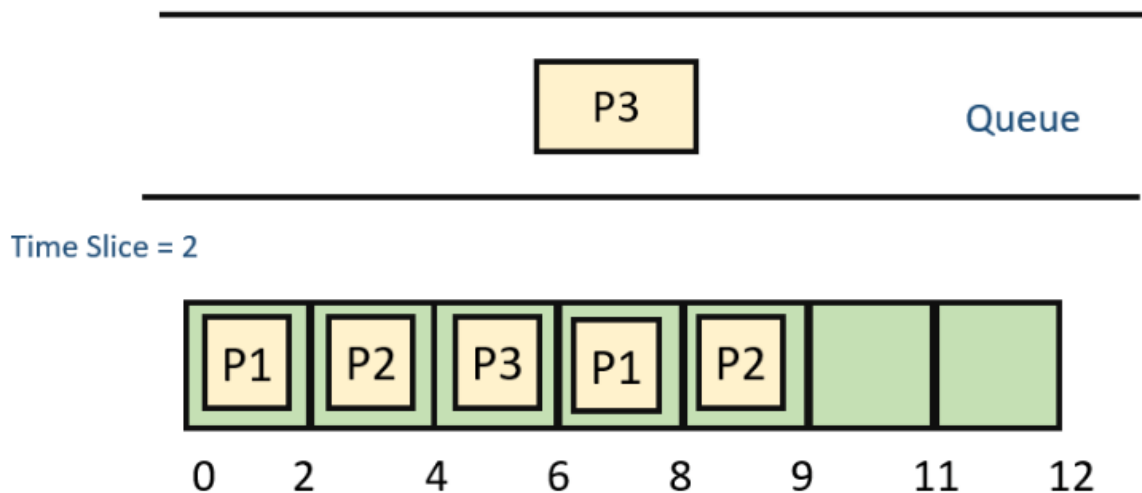


Рис. 2.19 П'ятий крок виконання Round Robin

6. P2 має час на виконання 3 мілісекунди. Він вже виконаний за 2 цикли. У момент 9 мілісекунд, P2 завершує виконання. Потім P3 починає виконання до завершення (оскільки лишився тільки процес P3 в черзі, P1 та P2 – виконані).

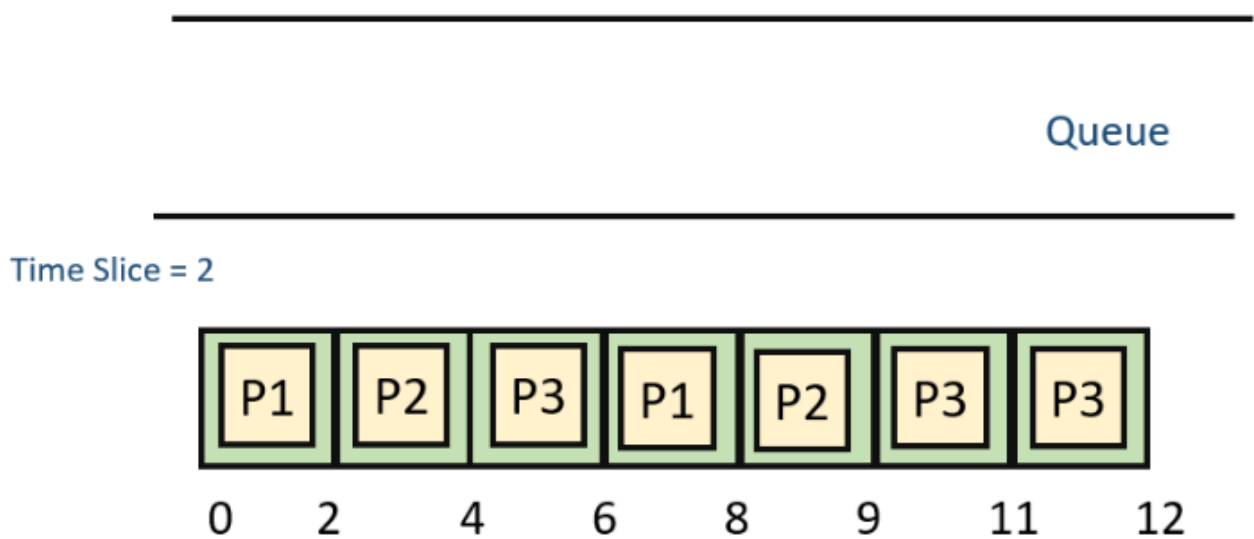


Рис. 2.20 Шостий крок виконання Round Robin

Згідно виконання методу Round Robin можна обчислити середній час очікування для наведеного прикладу:

1. Для процесу P1 (перший пріоритет на виконання) $= 0 + 4 = 4$ мілісекунд очікування на виконання.
2. Для процесу P2 (другий пріоритет на виконання) $= 2 + 4 = 6$ мілісекунд очікування на виконання.

3. Для процесу P3(третій пріоритет на виконання) = $4 + 3 = 7$ мілісекунд очікування на виконання.

Отже, Round Robin – це циклічне планування виконання процесів згідно заданого кванту часу. Цей метод не залежить від часу виконання будь-якого процесу, оскільки має фіксований час виконання для будь-якого процесу.

2.3.7 Weighted Round Robin

У багатьох корпоративних середовищах, де серверні вузли з різною характеристикою потужності та характеристикою обробки для розміщення служб та обробки кінцевих точок, часто доводиться розподіляти навантаження обробки на основі індивідуальних можливостей сервера, щоб деякі кінцеві точки не були несправедливо обтяжені запитами. Просте кругове або випадкове балансування навантаження не полегшує проблеми, але для вирішення цієї проблеми може використовуватися зважений балансувач кругового або зваженого випадкового навантаження. Weighted Round Robin[28] - це алгоритм планування, який використовується в мережах для планування потоків даних, але також використовується для планування процесів. Зважений циклічний алгоритм є узагальненням циклічного планування. Він обслуговує набір черг або завдань. У той час як циклічний перебір черг(завдань) дає одну можливість обслуговування за цикл, зважений циклічний перебір пропонує кожному процесу фіксовану кількість можливостей, робоча вага, встановлений в конфігурації. Потім це дозволяє впливати на долю ємності, що отримується кожної чергою / завданням. У комп'ютерних мережах можливість обслуговування - це передача одного пакету, якщо обрана черга не порожня. Якщо всі пакети даних мають однаковий розмір, WRR є найпростішою реалізацією розподілу навантаження в мережі.

Отже, сенс вдосконалення методу Round Robin полягає в наступному: кожного серверу присвоюється ваговий коефіцієнт відповідно до його продуктивності і потужності. Це допомагає розподіляти навантаження більш гнучко: сервери з великою вагою обробляють більше запитів. Однак усіх проблем з відмовостійкістю цей метод не вирішує.

2.3.8 Least Connections

Продуктивність мережі дуже важлива річ, це означає, що важливо, щоб інфраструктура в мережі підтримувала потребу в швидкості передавання даних. Алгоритми балансування навантаження є невід'ємною частиною підвищення продуктивності мережі і можуть одночасно покращувати - або погіршувати – продуктивність. Алгоритм балансування навантаження "Least connections"(найменша кількість зв'язків/підключень) використовує кількість поточних підключень до кожного екземпляру програми (члена) для прийняття рішення про збалансування навантаження. Тобто передумовою цього алгоритму є загальне припущення, що менша кількість з'єднань (і, отже, менша кількість користувачів) означає менше навантаження і, отже, кращу продуктивність. Якби всі робочі навантаження всередині програми вимагали однакових ресурсів це метод був би ефективним у розподілі навантаження в мережі, але, на жаль, на даний момент результатом є нерівномірний розподіл навантаження, що призводить до непередбачуваних коливань продуктивності у міру збільшення попиту. Наприклад, користувач, який використовує систему, відправляє принаймні один, якщо не більше, запитів до бази даних, щоб перевірити облікові дані, а потім оновити систему, щоб вказати діяльність. Залежно від характеру заявки, інші внутрішні програми вимагають різної кількості ресурсів. Деякі з них важкі запити до оперативної пам'яті, інші важкі запити до процесора, інші важкі запити до файлів або баз даних. Чим більша кількість користувачів в системі тим більша завантаженість системи, згідно відправлених користувачами запитів. В такому випадку система більш навантажена, і продуктивність може вже почати страждати. Однак, коли надходить наступний запит, балансувач навантаження, використовуючи алгоритм "найменших з'єднань", вибере останнього члена, збільшуючи навантаження на цього члена і, ймовірно, може виникнути подальше погіршення продуктивності мережі.

Отже, сенс алгоритму «найменшого зв'язку» полягає в тому, що екземпляр програми з найменшою кількістю зв'язків є найменш завантаженим, але для того щоб в цьому переконатись потрібно використовувати додаткове ПО для

моніторингу навантаження на сервер, а також збір інформації щодо використання процесора, пам'яті і т.д.

2.4 Переваги та недоліки проаналізованих методів розподілу навантаження в мереж

На таблиці 2.1 зображені переваги розподілу навантаження в мережі методів FCFS, Round Robin, Weighted Round Robin.

Таблиця 2.1. Переваги розподілу навантаження проаналізованих методів

Переваги		
FCFS	Round Robin	Weighted Round Robin
Найпростіша форма алгоритму планування процесора	Не залежить від часу відправлення/обробки пакета даних	Пропонує кожному процесу фіксовану кількість можливостей, робоча вага, встановлену в конфігурації
Легко програмувати	Має фіксований квант часу на виконання процесів	Розпізнає параметри кожного серверу
	Всі завдання отримують справедливий розподіл ЦП	

На таблиці 2.2 зображені недоліки розподілу навантаження в мережі методів FCFS, Round Robin, Weighted Round Robin.

Таблиця 2.2. Недоліки розподілу навантаження проаналізованих методів

Недоліки		
FCFS	Round Robin	Weighted Round Robin
Ніколи не звільнить ЦП, поки не завершить виконання виконуючого процесу.	Продуктивність сильно залежить від кванта часу	Витрачає багато часу на аналіз параметрів серверів
Середній час очікування високий	Неможливо встановити пріоритети для процесів	
Простий – не ефективний	Менший квант часу призводить до збільшення накладних витрат на переключення контексту в системі	
Неможливо встановити пріоритети для процесів	Витрачає багато часу на перемикання контексту	

Висновки:

1. Проаналізовані сучасні методи розподілу навантаження, їх принципи роботи та реалізація.
2. Приведені приклади виконання проаналізованих методів розподілу навантаження.
3. Сформовані переваги та недоліки проаналізованих методів розподілу навантаження в мережах

РОЗДІЛ 3

РОЗРОБКА МОДИФІКОВАНОГО МЕТОДУ РАЦІОНАЛЬНОГО РОЗПОДІЛУ НАВАНТАЖЕННЯ

У розділі описано розроблення модифікованого методу раціонального розподілу навантаження на основі проведеного аналізу з використанням методу розподілу навантаження Round Robin та висвітлено процес проведення імітаційного моделювання з метою перевірки роботи запропонованого підходу та проаналізовано результати моделювання.

3.1 Збір інформації

На етапі первинного збору інформації необхідно проаналізувати роботу мережі в кластері веб-серверів, а також на основі проведеного аналізу покращити метод розподілу навантаження в мережі.

Попит на веб-сервіси продовжує зростати, а системи веб-серверів стають більш напруженими, ніж будь-коли. Проблеми з продуктивністю веб-архітектури почали навіть погіршуватимуться через поширення різнотипових клієнтських пристроїв, необхідність аутентифікації клієнта та налаштування безпеки системи, підвищену складність проміжного програмного забезпечення та прикладного програмного забезпечення та високі вимоги до доступності корпоративних центрів обробки даних та електронних систем.

Інтернет перетворюється з інфраструктури спілкування та перегляду на носій для ведення особистого життя, бізнесу, комунікацій та фінансових операцій, що робить ефективність роботи кластерних веб-сервісів все більш актуальною. Веб-кластер має високу масштабованість і може легко покращити можливості обчислення та доступність мережі, додавши більше вузлів сервера. Саме при такій умові і був розроблений модифікований метод розподілу навантаження в мережі.

Розглянемо, наприклад, політику планування, яка називається політикою, орієнтованою на клієнта (CAP - client-aware policy), для веб-комутаторів, що працюють на рівні 7 стеку протоколів OSI. CAP класифікує запити клієнтів на основі їх впливу на основні серверні ресурси, тобто мережевий інтерфейс,

центральний процесор, диск. Під час виконання CAP планує запити клієнтів, що надходять до веб-кластера, з метою спільного використання всіх класів послуг між вузлами сервера. CAP приймає динамічні рішення, переглядаючи лише запити клієнтів, а не стан сервера. Крім того, CAP має перевагу, забезпечуючи надійні результати для дуже різних класів веб-служб, оскільки не вимагає детального налаштування системних параметрів, на відміну від того як це робить багато інших диспетчерських політик серверного програмного забезпечення. Також ефективною реалізацією балансування навантаження в кластері веб-серверів є алгоритм балансування навантаження, який класифікує запити на різні класи. При цьому такий алгоритм динамічно вибирає запит із класу і призначає запит серверу. Щоб уникнути ненадійного вимірюваного використання мережевих компонентів в умовах коливання навантажень, такий алгоритм балансування навантаження використовує модель організації обслуговування для прогнозування використання кожного сервера.

Саме на основі прогнозування навантаження на веб-сервер в кластері, був розроблений передбачуваний модифікований алгоритм балансування навантаження Round Robin. Сенс цього методу полягає у використанні величини навантаження попереднього моменту(виконаного процесу) для прогнозування поточного стану завантаження веб-серверів у кластері. Завдяки прогнозуванню, балансувач навантаження може приймати рішення щодо диспетчеризації, в той час коли вибраний веб-сервер обробляє запит користувача, це значно економить час. Тобто завдяки цьому методу можна приблизно дізнатись робочий стан, споживання ємності та допустиме навантаження веб-серверів, і в цьому випадку можна навіть отримати загальну оцінку вартості та продуктивності веб-кластера, що робить цей метод актуальним.

3.2 Архітектура кластерного веб-серверу

Кластерна веб-система складається з двох або більше серверних машин, які розміщені разом в одному місці і підключені через високошвидкісну мережу для вирішення запитів користувачів. Кожна серверна машина діє як вузол, який має власний диск та операційну систему. Хоча кластер має велику кількість веб-

серверів, він використовує лише одне ім'я хосту та одну віртуальну IP-адресу, щоб забезпечити єдиний інтерфейс для зовнішніх користувачів та мати механізм, який контролює всі запити сайту та виконує розподіл послуг серед усіх серверів. Одним з найважливіших компонентів є балансування навантаження, механізм, який спрямовує навантаження і відправляє запити клієнтів на всі вузли сервера. Алгоритм балансування навантаження знаходиться в балансувачі навантаження для виконання логіки, яка контролює призначення запиту. Вузол сервера, який отримує запит користувача, може обробити його самостійно. Балансувач навантаження може достатньо рівномірно розподіляти запити клієнтів до вузлів сервера, достатньо швидко надавати важливі дані та вирішувати проблеми обслуговування одночасних відвідувань кількох користувачів. Ця технологія настільки економна, що завдяки її застосуванню можемо отримати мінімальну продуктивність ЕОМ з мінімальними вкладеннями.

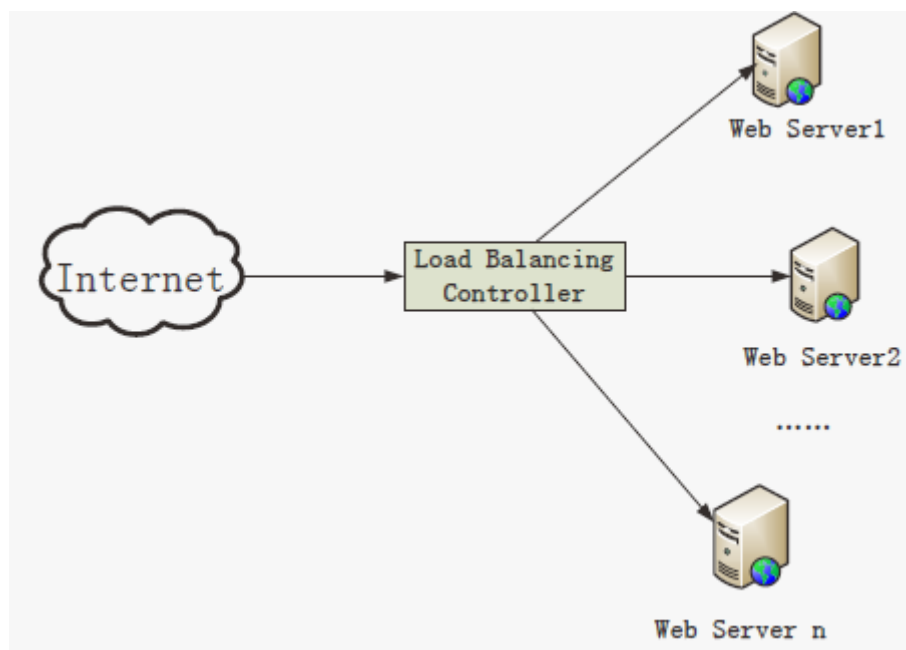


Рис. 3.1 Архітектура кластеру веб-серверів

Рівень доступу спрямовує запити до кластера та виконує аутентифікацію та основні функції брандмауера, такі як виявлення вторгнень до мережі. Також статичний вміст можна безпосередньо подавати через вузли веб-кешу. Веб-комутатор можна класифікувати на два типи, веб-комутатори рівня 4 і рівня 7, відповідно до рівня стеку протоколів OSI, яким вони керуються. Вони досить

різні, оскільки веб-комутатори рівня-4 використовують розподіл інформації про вміст, тоді як веб-комутатори рівня-7 розгортають інформаційно розподілений вмісту інформації. Тобто веб-комутатор рівня-4 працює на рівні TCP / IP і повинен підтримувати таблицю прив'язки, щоб асоціювати кожен сеанс TCP-клієнта з цільовим сервером, тоді як комутатор рівня-7 може встановити повне з'єднання TCP з клієнтом і перевірити вміст HTTP-запиту до прийняття рішення про відправлення.

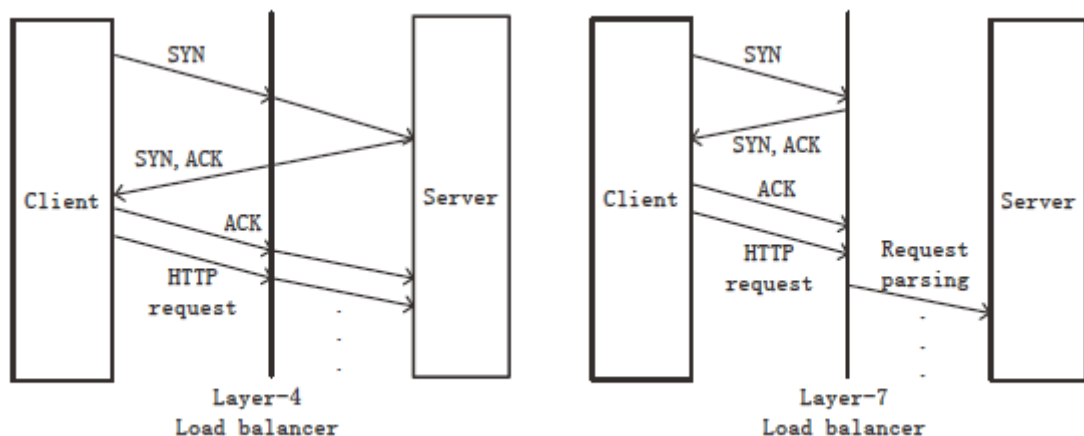


Рис. 3.2 Робота маршрутизації рівня 4(зліва) та рівня 7(справа)

Алгоритм балансування навантаження, що працює в балансувачі навантаження, має велике значення для підвищення продуктивності кластера. Коли надходить новий запит, він вибирає найбільш відповідний сервер і призначає йому запит. Алгоритми балансування навантаження працюють за принципом того, в якій саме ситуації призначається навантаження, під час компіляції чи під час виконання.

3.3 Модифікований метод розподілу навантаження

На відміну від загального алгоритму Round Robin, модифікація алгоритму кругового планування враховує інформацію про стан системи для оптимізації результатів диспетчеризації. Крім того, модифікований Round-robin використовує круговий список та вказівник на останній вибраний сервер для прийняття рішень щодо відправки.

В даний час в більшості досліджень в балансувачі навантаження використовуються деякі алгоритми, встановлені заздалегідь. Коли надходить

новий запит користувача, він працює, а потім приймає диспетчерські рішення, які рідко беруть до уваги прогнозування розподілу навантаження. Тому, в описі запропонованого методу розглядається більше процес про прогнозування балансу навантаження. Відповідно до результату прогнозування балансу навантаження можна оцінити поточну ситуацію всіх серверів у кластері, а диспетчер може прийняти відповідні алгоритми балансування навантаження та внести певні корективи для досягнення кращого балансу навантаження. Завдяки прогнозам, балансувач навантаження може приймати рішення щодо диспетчеризації, поки обраний веб-сервер обробляє запит користувача, що економить час. А для великомасштабної мережі, ведучи облік навантаження на кожен серверний вузол, можна приблизно знати робочий стан, споживання ємності та величину допустимого навантаження веб-серверів, і в цьому випадку є можливість навіть отримати загальну оцінку вартості та продуктивності веб-кластера.

Запропонований алгоритм будується на наступних припущеннях:

1. Стан завантаження за минулий час все ще впливає на поточний стан завантаження. Кожен балансувач навантаження у веб-кластері щодня веде запис кількості завантаження (consolidation).
2. Вплив кількості завантажень минулого часу на поточний стан веб-сервера змінюється щомиті. Чим ближче діапазон часу, тим більший вплив він робить. Для опису ваги завантаженості використовуються різні фактори прогнозуючої цінності. Дуже важливо вибрати відповідне значення для прогнозуючих факторів.
3. Оскільки балансувач навантаження не може щомиті реєструвати кількість навантаження на вузлах сервера, слід вибрати кілька останніх величин навантаження, щоб передбачити поточний стан навантаження, це допоможе уникнути занадто великих обчислень.

Припустимо, що у веб-кластері є N веб-серверів, від x_1 до x_N . Виходячи з наведених вище припущень, виберемо останню кількість завантажень p кожного запису завантаження веб-сервера, щоб робити прогнози, від x_1 до x_p . x_i - означає

значення прогнозуючого навантаження серверного вузла i , $x_{i,j}$ - вказує j^{st} - коефіцієнт прогнозуючого значення для прогнозуючого навантаження вузла сервера i α_j - означає вагу j^{st} коефіцієнта прогнозуючої цінності. x_{max} - означає максимальне значення від x_i та x_{min} - означає мінімальне значення x_i . Отже, маємо $x'_1 = \alpha_1 x_{i,1} + \alpha_2 x_{i,2} + \dots + \alpha_j x_{i,j} + \dots + \alpha_p x_{i,p} (\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_p)$

Спочатку отримуємо останній обраний вузол сервера i . Використовуючи планування Round-Robin, цикл повинен надіслати новий запит на наступний вузол сервера $i+1$. За допомогою модифікованого методу планування Round-Robin можна спочатку розрахувати суму навантаження всіх веб-серверів. Після обробки процесу S_i подає запит на серверний вузол S_{i+1} , якщо сума навантаження серверного вузла S_{i+1} не є найбільшою. Припустимо що на даний момент надходить наступний новий запит. Наприклад, у момент t сума навантаження S_{i+1} не є найбільшою, тому далі обробляється процес S_{i+1} . Однак, коли в момент $t + 1$ надходить ще один новий запит, сума навантаження S_{i+1} стає найбільшою випадково, оскільки процес на останньому вузлі ще міг не закінчитись і в такому випадку новий запит стає в чергу. Отже, рішення про диспетчеризацію, прийняте на момент t , більше не підходить і веде до дизбалансу навантаження. При використанні прогнозуючого значення суми навантаження для прийняття диспетчерських рішень, можна досягти кращого балансу навантаження. Коли запит надходить на балансувач навантаження, спочатку отримується останній обраний вузол S_i . Після потрібно отримати останній статус завантаження p усіх вузлів сервера, від x_1 до x_p . Також потрібно вибрати відповідний прогнозуючий фактор відповідно, з α_i до α_p . Значення факторів прогнозування є суттєво важливими, тому що воно впливає безпосередньо на ефект прогнозування і дуже важливо вибрати відповідне значення прогнозуючого фактора в алгоритмі. Виходячи з цього, модифікований метод циклічного методу розподілу навантаження обчислює прогнозуючу суму навантаження кожного сервера, якщо передбачена сума навантаження S_{i+1} не найбільша, новий запит призначається S_{i+1} . Однак, якщо передбачувана сума

навантаження S_{i+1} найбільша по вазі завантаженості, новий запит призначається S_{i+2} для досягнення балансу навантаження. Робота алгоритму показана на блок-схемі прогнозуючого модифікованого циклічного планування навантаженням.

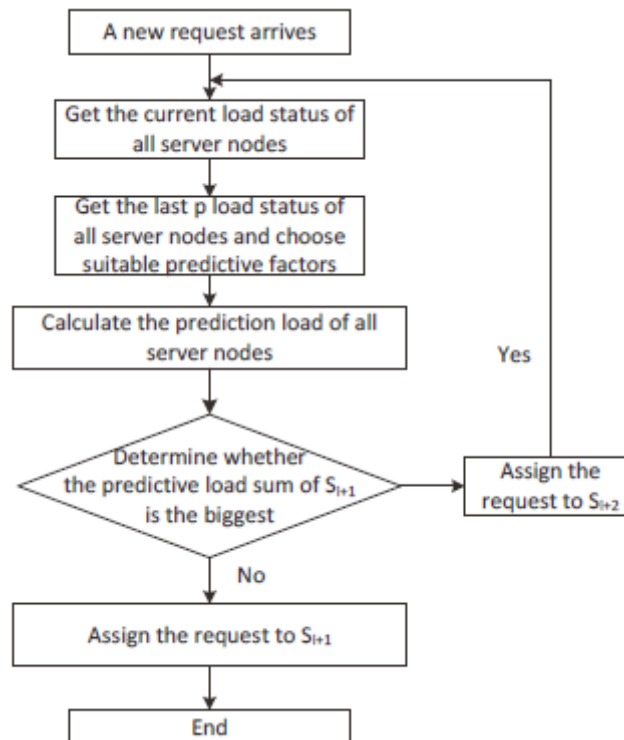


Рис. 3.3 Блок схема модифікованого методу циклічного планування навантаження

3.4 Експериментальне дослідження ефективності запропонованого методу розподілу навантаження

Було проведено експериментальне дослідження в MATLAB ефективності запропонованого методу розподілу навантаження. Маючи 4 сервери була проведена генерація 100 різних випадкових навантажень, дослідження були проведені з використанням звичайного алгоритму циклічного планування навантаження, модифікованого методу розподілу навантаження та прогнозуючого завантаженості модифікованого методу розподілу навантаження на основі циклічного планування навантаженням Round Robin.

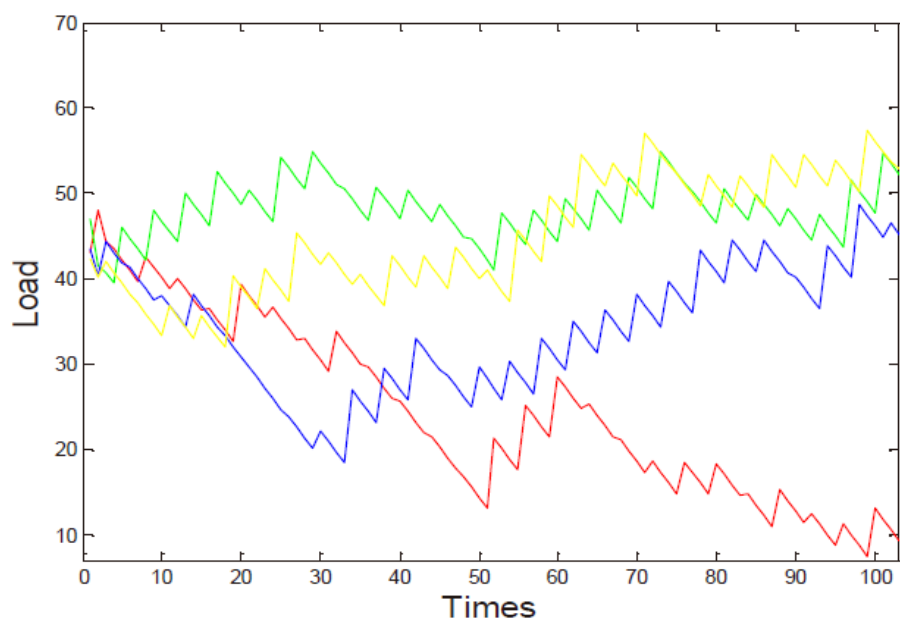


Рис.3.4 Завантаження кластера веб-серверів за допомогою планування Round-Robin

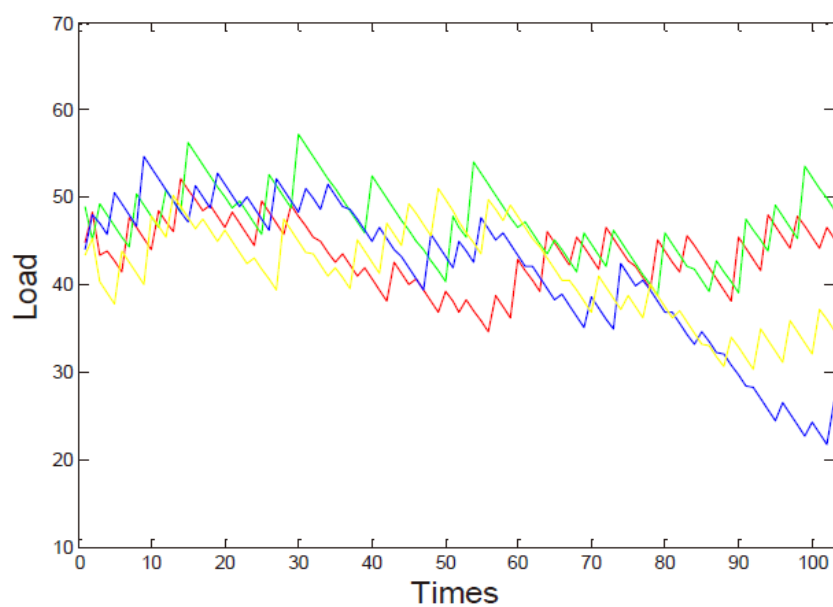


Рис. 3.5 Завантаження кластера веб-серверів за допомогою модифікованого методу планування Round-Robin

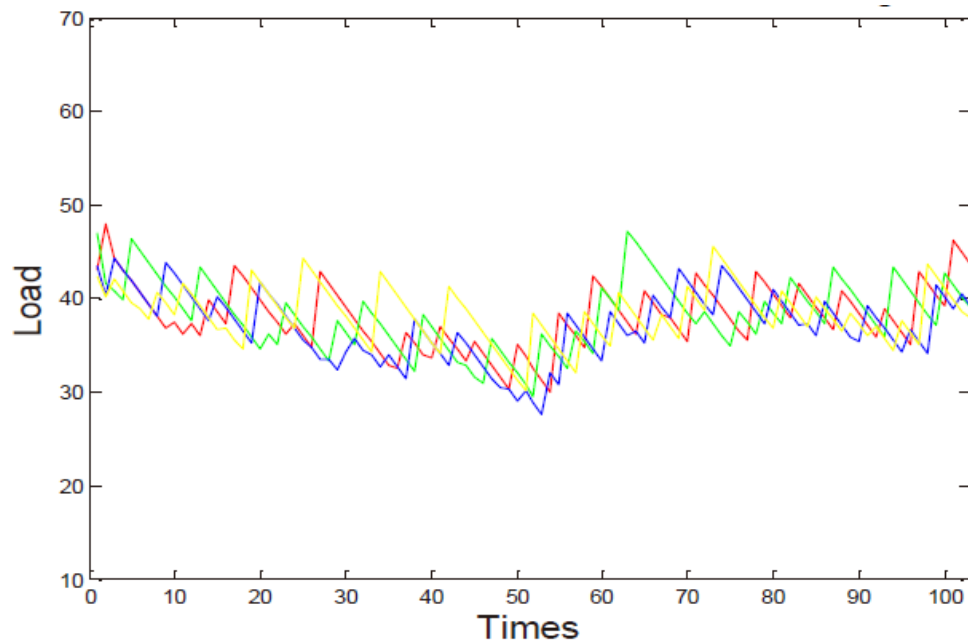


Рис. 3.6 Завантаження кластера веб-серверів за допомогою модифікованого методу з прогнозуванням планування навантаження Round-Robin

Також були розраховані діапазон навантаження та дисперсія навантаження чотирьох веб-серверів, що розгортають три різні політики планування розподілу навантаження в кластері між веб-серверами(таблиця 3.1).

Таблиця 3.1

Результати трьох досліджуваних політик планування навантаження

Політика планування навантаження	Round Robin	Модифікований Round Robin	Модифікований Round Robin з прогнозуванням навантаження
Діапазон навантаження	55.16	40.86	17.26
Дисперсія навантаження	104.47	102.53	9.78

Моделювання згідно діапазону та дисперсії навантаження досліджуваних методі зображені на рис 3.7-3.8.

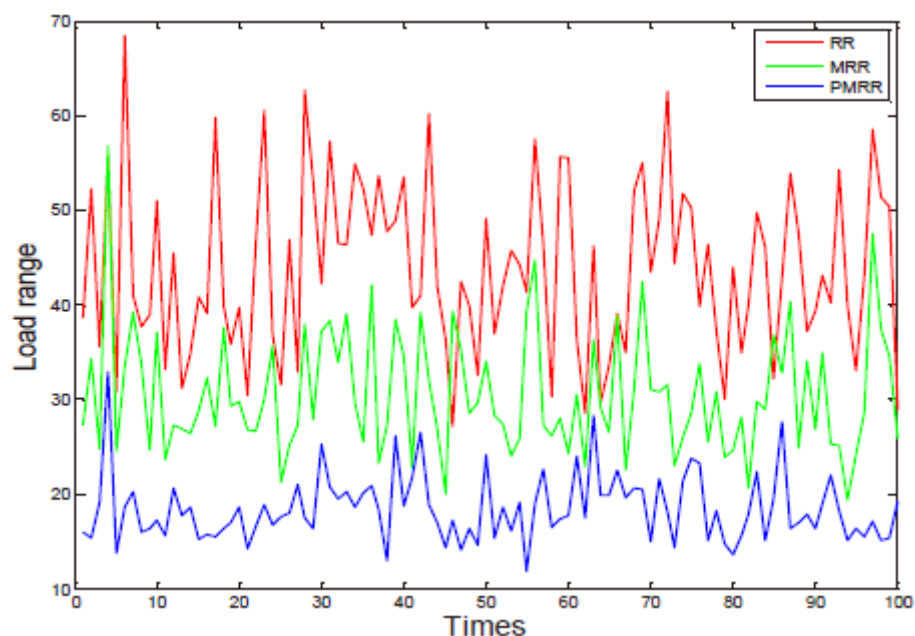


Рис. 3.7 Діапазон навантаження з трьох різних алгоритмів

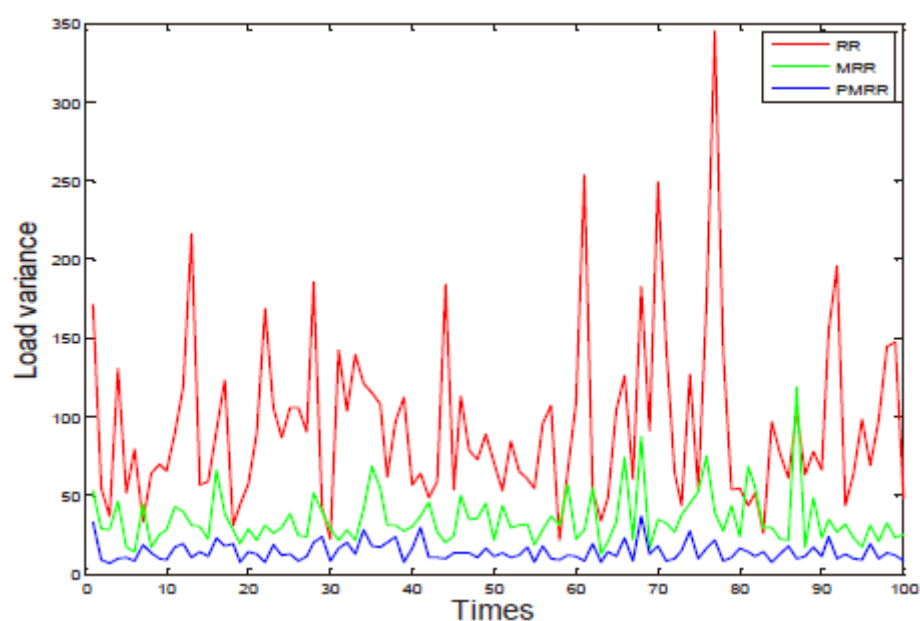


Рис. 3.8 Дисперсія навантаження з трьох різних алгоритмів

Висновки:

1. Побудовано блок-схему виконання модифікованого методу розподілу навантаження за рахунок прогнозування навантажень на основі методу Round Robin;

2. Запропонований метод порівняно з циклічним плануванням задач Round Robin;

3. В процесі порівняння цих підходів виявлено, що запропонований метод, згідно експериментального дослідження зменшує дисперію та діапазон навантаження на відміну від звичайного методу циклічного планування навантаження Round Robin.

РОЗДІЛ 4

РОЗРОБКА СТАРТАП-ПРОЕКТУ

У розділі проводиться аналіз стартап-проекту за темою даної роботи для визначення принципової можливості його ринкового впровадження та можливих напрямів реалізації цього впровадження.

4.1 Опис ідеї стартап-проекту

У табл. 4.1 наведено зміст ідеї стартап-проекту, в межах якої потрібно знайти групу потенційних клієнтів, напрямки, де можна реалізувати проект та вимоги для користувача. Визначимо потенційну цільову аудиторію проекту, основних конкурентів на ринку та конкурентну перевагу запропонованого продукту.

Таблиця 4.1
Опис ідеї стартап-проекту

Зміст ідеї	Напрямок застосування	Переваги для користувача
Модифікація методу розподілу навантаження Round Robin	Інфокомунікаційні мережі та мережі нового покоління	Ефективне використання мережевих ресурсів в мережі
		Розподіл навантаження на основі прогнозуючого завантаження серверів
		Можливість отримати робочого стану, споживання ємності та навантаження веб-серверів.

Аналіз потенційних техніко-економічних переваг ідеї (відмінність від існуючих методів) порівняно із пропозиціями конкурентів передбачає:

1. Визначення переліку техніко-економічних властивостей та характеристик;
2. Визначення попереднього кола конкурентів (продуктів-конкурентів) або товарів-аналогів, що вже існують на ринку, та збір інформації щодо значень техніко-економічних показників для ідеї власного проекту та проектів-конкурентів відповідно до визначеного вище переліку;

3. Порівняльний аналіз показників: для власної ідеї визначаються показники, що мають: а) гірші значення (W, слабкі); б) аналогічні (N, нейтральні) значення; в) кращі значення (S, сильні) (табл. 4.2).

Таблиця 4.2

Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№	Техніко-економічні характеристики ідеї	(Потенційні) послуги/концепції конкурентів			W	N	S
		Мій проект	FIFO	Round Robin			
1	Призначення	Використання в мережах зв'язку	Використання в мережах зв'язку	Використання в невеликих мережах зв'язку			+
2	Надійності	Довговічність використання	Довговічність використання	Довговічність використання		+	
3	Технологічні	Збільшена відмовостійкість мережі та покращена ефективність роботи мережі	Легко програмується	Має квант часу, за який виконується інформаційний процес в мережі			+
4	Ергономічні	Метод зрозумілий в програмуванні	Метод зрозумілий в користуванні та налаштуванні	Метод зрозумілий в користуванні та налаштуванні		+	
5	Безпеки	Безпечно	Безпечно	Безпечно		+	

4.2 Технологічний аудит ідеї проекту

В межах даного підрозділу проведено аудит технології, за допомогою якої можна реалізувати ідею проекту (технології створення продукту). Визначення технологічної цінності ідеї проекту передбачає аналіз таких складових (табл.4.3):

1. технологія, за якою буде виготовлено товар згідно ідеї проекту;
2. аналіз наявності такої технологій;
3. доступність технологій автору проекту.

Таблиця 4.3

Технологічна здійсненність ідеї проекту

№	Ідея проекту	Технології реалізації	Наявність технології	Доступність технологій
1	Математичне дослідження характеристики мережі	Програмне середовище MATLAB	+	–
2	Модифікація методу розподілу навантаження в мережі	ns-3 simulator	+	+
3	Експериментальне дослідження роботи запропонованого методу	ns-3 simulator	+	+
Обрана технологія для аналізу та дослідження ідеї проекту: Програмне середовище ns-3 simulator				

4.3 Аналіз ринкових можливостей запуску стартап-проекту

Визначення ринкових можливостей, які можна використовувати під час ринкового впровадження проекту, та ринкових завад, які можуть перешкодити реалізації проекту, дає можливість спланувати напрямки розвитку проекту з врахуванням стану ринкового середовища, потреб потенційних клієнтів та пропозицій проектів-конкурентів.

Спочатку проведено аналіз попиту: наявність попиту, обсяг, динаміка розвитку ринку (табл. 4.4).

Таблиця 4.4

Попередня характеристика потенційного ринку стартап-проекту

№	Показники стану ринку (найменування)	Характеристика
1	Кількість головних гравців, од	4
2	Динаміка ринку (якісна оцінка)	Зростає
3	Наявність обмежень для входу	Немає
4	Специфічні вимоги до стандартизації сертифікації	Немає
5	Середня норма рентабельності в галузі ринку, %	200

За результатами попереднього оцінювання ринок є привабливим для входження. Потенційні групи клієнтів, їх характеристики, орієнтовний перелік

вимог до товару для кожної групи наведено в таблиці 4.5. Проведено аналіз ринкового середовища: складені таблиці факторів, що сприяють ринковому впровадженню проекту, та факторів, що йому перешкоджають (табл.4.6-4.7). Фактори в таблицях подані в порядку зменшення значущості.

Таблиця 4.5

Характеристика потенційних клієнтів стартап-проекту

Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до послуги
Відмовостійка робота мережі	Інфокомунікаційні мережі	Поведінку клієнта Формують потреби; особливостей купівлі та експлуатації продукту немає	Мережа має стабільно забезпечувати доступ до надаючих послуг

Таблиця 4.6

Фактори загроз

№	Фактор	Зміст загрози	Можлива реакція компанії
1	Нестача кваліфікованих кадрів	Для налаштування мережових компонентів потрібна команда спеціалістів	Пошук персоналу у науково-дослідних інститутах та коледжах
2	Нестача ресурсів	Для створення продукту потрібно технічне обладнання та програмне забезпечення для тестування працездатності мережі	Укладання договорів з державними або приватними структурами для фінансування та надання можливостей для налаштування мережі
3	Фінансова нестабільність	Необхідність в коштах для налаштування та забезпечення якості обслуговування мережі.	Пошук інвесторів

Таблиця 4.7

Фактори можливостей

№	Фактор	Зміст можливості	Можлива реакція компанії
1	Попит	Існування постійного попиту на ефективні методи розподілу навантаження означає, що більшість клієнтів зацікавлені у оновленні своїх програмних компонентів в мережі.	Налаштування реклами та її розвиток, просування у Інтернеті, участь у конференціях.
2	Науково технічний прогрес	Постійний розвиток інформаційних технологій та збільшення програмних алгоритмів дозволяють доцільно вирішити поставлені завдання	Ознайомлення з останніми інноваціями та трендами у відповідній галузі; відвідування наукових конференцій.

Проведений аналіз пропозиції: визначені загальні показники конкуренції на ринку (табл. 4.8). Після аналізу конкуренції проведено більш детальний аналіз умов конкуренції в галузі (за М. Портером) (табл. 4.9). Сильні позиції компанії за кожним з факторів М. Портера означають і її можливості забезпечити необхідні кроки обороту капіталу та здатність впливати на інших агентів ринку, надаючи їм власні умови співпраці. Характеристики факторів моделі відрізняються для різних галузей та змінюються з часом. На основі аналізу конкуренції, наведеного в табл. 4.9, а також із урахуванням характеристик ідеї проекту (табл. 4.2), вимог споживачів до товару (табл. 4.5) та факторів маркетингового середовища (табл. 4.6-4.7) визначається та обґрунтовується перелік факторів конкурентної спроможності. Аналіз наведено в табл. 4.10.

Таблиця 4.8

Ступеневий аналіз конкуренції на ринку

№	Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1	Тип конкуренції: монополія	На ринку в межах України існує 2 подібні компанії, але вони налаштовують свою мережу за своїми стандартами	Підвищувати якість послуги (досягати кращих показників) за рахунок використання передових технологій та ефективних програмних алгоритмів
2	За рівнем конкурентної боротьби: державний	Компанії-конкуренти що знаходяться в Україні, поки що не оприлюднюють технологію, яку використовують.	Налаштувати мережу, що відповідатиме міжнародним стандартам
3	За галузевою ознакою: внутрішньогалузева	Економічна боротьба між наданими послугами, які діють в одній галузі економіки, виробляють і реалізують однакові послуги, що задовольняють одну й ту саму потребу, але мають відмінності у виробничих затратах, якості та ціні	Слідкувати за розвитком продуктів у конкурентів
4	Конкуренція за видами послуг: товарно-видова	Конкуренція між послугами одного виду	Покращувати рівень якості надання послуги
5	За характером конкурентних переваг: цінова	Передбачає надання послуг за нижчими цінами ніж у конкурентів	Надавати послугу за меншою ціною
6	За інтенсивністю: марочна	В сучасній економічній ситуації боротьба носить явно виражений марочний характер, велике значення набуває брендинг	Реклама наданих послуг, створення символіки продукту

Таблиця 4.9

Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти
	HostPro, Hostinger	Немає	Немає	Вимоги до якості
Висновки	Інтенсивність конкурентної боротьби висока	Можливість виходу на ринок є	Постачальники не ставлять свої умови роботи на ринку	Послуга має задовольняти вимоги клієнтів

Таблиця 4.10

Обґрунтування факторів конкурентоспроможності

№	Фактор конкурентоспроможності	Обґрунтування (чинники, що роблять фактор для порівняння конкурентних проектів значущим)
1	Ступінь задоволення потреб користувача	Продукт має забезпечувати передавання даних зі швидкістю яка задовольняє користувача та доступ до сервісів, які потрібні користувачу
2	Сумісність з існуючими системами	Продукт не має створювати завад для інших мереж зв'язку
3	Якість розробки з точки зору показників надійності	Продукт має забезпечувати надійність переданих даних
4	Наявність кваліфікованих кадрів в команді	Науковці з досвідом та високим рівнем підготовки, які зацікавлені в розподілу навантаження в мережі
5	Економічний	Ціна товару має відповідати якості надання послуг

На фінальному етапі ринкового аналізу можливостей впровадження проекту виконано SWOT-аналіз (матриця аналізу сильних (Strength) та слабких (Weak) сторін, загроз (Troubles) та можливостей (Opportunities) (табл. 4.11) на основі виділених ринкових загроз та можливостей, сильних і слабких сторін. Перелік ринкових перешкод та ринкових можливостей складено на основі аналізу факторів перешкод та факторів можливостей маркетингового середовища. Ринкові перешкоди та ринкові можливості є наслідками впливу факторів, і, на відміну від них, ще не є реалізованими на ринку та мають певну ймовірність здійснення.

На основі SWOT-аналізу розроблено альтернативи ринкової поведінки (перелік заходів) для виведення стартап-проекту на ринок та орієнтовний час їх

ринкової реалізації з огляду на потенційні проекти конкурентів, що можуть бути виведені на ринок. Визначені альтернативи аналізуються з точки зору термінів та ймовірності отримання ресурсів (табл. 4.12).

Таблиця 4.11

SWOT-аналіз стартап-проекту

Сильні сторони	1. Економічна (ціна товару повинна відповідати якості надаючих послуг). 2. Ступінь задоволення потреб користувача. 3. Якість розробки з точки зору факторів надійності
Слабкі сторони	1. Недостатність наукових та технічних ресурсів. 2. Присутність конкурентів на міжнародному ринку 3. Популярність брендингу. 4. Нестабільна позиція на ринку.
Можливості	1. Мінімізування витрат на виробництво продукції. 2. Покращення якості продукту. 3. Формування попиту на товар за рахунок просування реклами. 4. Освоєння міжнародного ринку. 5. Залучення нових користувачів. 6. Розширення складу команди. 7. Нові технології та методи.
Загрози	1. Несприятлива економічна ситуація в країні. 2. Додатковий державний контроль якості продукту. 3. Товари-аналоги.

Таблиця 4.12

Альтернативи ринкового впровадження стартап-проекту

№	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1	Пошук наукових та технічних ресурсів, пошук інвесторів, поглиблене дослідження, моделювання та тестування, створення та просування реклами.	70%	3 роки
2	Налагодження виробничого процесу, дослідження потреб та вимог до продукту споживачів, пошук коштів, пошук технологічних компонентів, пошук обладнання, створення та налаштування програмного забезпечення та його подальше тестування.	65%	4 роки

Із зазначених альтернатив обрано альтернативу № 1, так як для неї отримання ресурсів є більш ймовірним, а строки реалізації – більш стислими

4.4 Розроблення ринкової стратегії проекту

Розроблення ринкової стратегії першим ділом передбачає визначення стратегії охоплення ринку: опис цільових груп потенційних споживачів, що наведено в табл. 4.13.

Таблиця 4.13

Вибір цільових груп потенційних споживачів

№	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Державні установи	Готові	Високий	Середня	Середня
2	Приватні підприємства	Готові	Високий	Висока	Середня
Які цільові групи обрано: обрано цільову групу №2.					

Для роботи в обраних сегментах ринку сформована базова стратегія розвитку (табл. 4.14). Вибір стратегії конкурентної поведінки наведено в табл. 4.15.

Таблиця 4.14

Визначення базової стратегії розвитку

№	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
1	Задоволення потреб вибраного цільового сегменту	Формування попиту у користувачів за рахунок унікальних характеристик та високої якості обслуговування продукту	Стратегія спеціалізації

Таблиця 4.15

Визначення базової стратегії конкурентної поведінки

№	Чи є проект унікальним на ринку?	Чи буде компанія шукати нових користувачів?	Чи буде компанія копіювати характеристики наданих послуг конкурента, і які?	Стратегія конкурентної поведінки
1	Так	Так	Ні	Стратегія лідера

На основі вимог споживачів з обраних сегментів до постачальника (стартап компанії) та до продукту (табл. 4.5), а також в залежності від обраної базової стратегії розвитку (табл. 4.14) та стратегії конкурентної поведінки (табл. 4.15) розроблено стратегію позиціонування, наведену в табл. 4.16. Стратегія позиціонування полягає у формуванні комплексу асоціацій, за яким споживачі мають ідентифікувати торгівельну марку/проект.

Таблиця 4.16

Визначення стратегії позиціонування

№	Вимоги до послуги цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту
1	Товар має забезпечувати швидкість передавання та забезпечувати якість наданих послуг	Стратегія спеціалізації	Високий ступінь задоволення потреб користувача, ефективність, нові технології та методи балансування, помірна ціна	Доцільне співвідношення ціна/якість, наукоємність, співпраця.

4.5 Розроблення маркетингової програми стартап-проекту

Першим кроком є формування маркетингової концепції товару, який отримає споживач. Для цього у табл. 4.17 наведені результати попереднього аналізу конкурентоспроможності товару.

Таблиця 4.17

Визначення ключових переваг концепції потенційного товару

№	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами
1	Рациональний розподіл навантаження та ефективне використання технологічних компонентів	Підвищення відмовостійкості та ефективності роботи мережі	Значне покращення роботи мережі
2	Якість	Стабільність роботи продукту в мережі	Використання високоякісного обладнання

Розроблено трьохрівневу маркетингову модель товару: детальна ідея продукту (та/або послуги), його фізичні складові, особливості процесу його надання (табл. 4.18).

Таблиця 4.18

Опис трьох рівнів моделі товару

Рівні послуги	Сутність та складові		
I. Послуга за задумом	Розподіл навантаження в мережі зв'язку		
II. Послуга у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх/Тл/ Е/Ор
	1. Забезпечення раціонального розподілу навантаження в мережі. 2. Помірна ціна. 3. Забезпечення якості обслуговування наданих послуг 4. Покращення відмовостійкості мережі		
II. Послуга із підкріпленням	До продажу: гарантія		
	Після продажу: встановлення, налаштування, обслуговування, підтримка.		
За рахунок чого потенційний товар буде захищено від копіювання: захист інтелектуальної власності			

Наступним кроком визначено цінові межі, якими необхідно керуватись при встановленні ціни на потенційний продукт (остаточне визначення ціни відбувається під час фінансово-економічного аналізу проекту), яке передбачає аналіз ціни на продукти-аналоги або продукти субституту, а також аналіз рівня доходів цільової групи споживачів (табл. 4.19). Аналіз проведено експертним методом.

Таблиця 4.19

Визначення меж встановлення ціни

№	Рівень цін на товари замітники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
1	2000 - 4000 у.о	5000 у. о	Високий	8000 – 11000 у.о

Наступним кроком визначено оптимальну системи збуту, в межах якої приймається рішення (таблиця 4.20):

1. збут власними силами або із залученням сторонніх посередників (власна або залучена система збуту);
2. вибір та обґрунтування оптимальної глибини каналу збуту;
3. вибір та обґрунтування виду посередників.

Останньою складовою маркетингової програми є розроблення концепції маркетингових комунікацій, що спирається на попередньо обрану основу для позиціонування, визначену специфіку поведінки клієнтів (табл. 4.21). Маркетингова комунікація охоплює всі діяльності підприємства, які спрямовані на інформування, переконання, нагадування споживачам та ринку в цілому про свої продукти та свою діяльність.

Таблиця 4.20

Формування системи збуту

№	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник послуги	Глибина каналу збуту	Оптимальна система збуту
1	Замовлення послуг на сайті або безпосередньо у розробників	Встановлення налаштування, тестування, зберігання	Канал нульового рівня	Виробник пропонує продукцію покупцям

Таблиця 4.21

Концепція маркетингових комунікацій

Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
Використання послуги для передавання даних без втрат та очікувань. Використання у комерційних цілях	Прямі канали комунікації, в яких дані передаються від інформатора до інформованої особи	Раціональний розподіл навантаження в мережі.	Поширення про новий метод модуляції. Поширення знань про продукт	Особлива увага приділяється розподілу навантаження в мережі. Відмінність між послугою і послугами замінниками.

Висновки

1. За результатами попереднього оцінювання ринок є привабливим для входження. Спостерігається попит на використання удосконаленого методу розподілу навантаження в мережі, що забезпечують підвищення відмовостійкості та ефективності роботи мережі.

2. Продукт є унікальним в своїй галузі. Конкурентів на міжнародному ринку невелика кількість. Для подальшого дослідження, розробки та тестування товару потрібно формувати команду спеціалістів та залучати інвесторів.

3. Для ринкової реалізації проекту доцільно обрати альтернативу – пошук наукових та технічних компонентів, залучення інвесторів, створення та просування реклами, взаємодія з користувачами.

ЗАГАЛЬНІ ВИСНОВКИ ПО РОБОТІ

У цій роботі запропонований модифікований метод розподілу навантаження в мережі на основі прогнозування завантаження серверів, що дозволяє раціонально розподіляти навантаження в мережі та підвищувати її відмовостійкість та ефективність роботи.

В роботі було отримано такі теоретичні результати:

1. Був проведений огляд можливостей інфокомунікаційних мереж, їх структура та склад основних технологічних компонентів, а також були розглянуті принципи побудови та роботи інфокомунікаційних мереж.

2. Були виділені основні причини перенавантаження в інфокомунікаційних мережах при використанні різних комутацій передачі даних та сформовані основні цілі для розподілу навантаження в інфокомунікаційних мережах.

3. Були проаналізовані методи розподілу навантаження в мережах, а також виділені їх переваги та недоліки.

4. Запропоновано модифікований метод розподілу навантаження в інфокомунікаційних мережах при прогнозуванні завантаження серверів у кластері на основі проведеного аналізу.

5. Був проведений експериментальний аналіз роботи запропонованого методу та підтвердження його удосконаленої роботи в порівнянні з розглянутим методом розподілу навантаження Round Robin.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Van Steen M. A brief introduction to distributed systems / M. van Steen. // Computing. – 2016. – №98. – С. 967–1009.
2. Таненбаум Э. Распределенные системы. Принципы и парадигмы / Э. Таненбаум, М. Ван Стеен., 2003. – 877 с. – («Классика Computer Science»). — ISBN 5–272–00053–6.
3. Boemer J. C. Dynamic models for transient stability analysis of transmission and distribution systems with distributed generation: An overview / J.C. Boemer, M. Gibescu, W.L. Kling // IEEE Bucharest PowerTech, Bucharest, 2009, pp. 1-8. doi: 10.1109/PTC.2009.5282177
4. Blair G. S. Building heterogeneous distributed multimedia systems / G. S. Blair // IEEE Colloquium on Building Distributed Systems, London, UK, 1990, pp. 611-613.
5. Austin M.A. Parallel distributed real-time systems in manufacturing (an aerospace view) / M. A. Austin // Proceedings of 5th International Workshop on Parallel and Distributed Real-Time Systems and 3rd Workshop on Object-Oriented Real-Time Systems, Geneva, Switzerland, 1997, pp. 284-288. doi: 10.1109/WPDRTS.1997.637992
6. Data Center Market and Technology Trends Power Electronics presentation [Електронний ресурс] // Yole Développement, APEC. – 2016. – Режим доступу до ресурсу: https://www.slideshare.net/Yole_Developpement/data-center-market-and-technology-trends-power-electronics-presentation-held-at-apec-2016-from-yole-dveloppement.
7. Бойчук Н. Я. Проблеми та перспективи розвитку національної економіки / Н. Я. Бойчук, Острянюк М.М. // Сучасні проблеми економіки та підприємництва. – 2017. – №19.
8. Barroso L. A. The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines / L. A. Barroso, J. Clidaras, U. Hoesle // USA: Morgan & Claypool 2nd ed. – 2013.

9. Dargie W. Analysis of the Power and Hardware Resource Consumption of Servers Under Different Load Balancing Policies / W. Dargie and A. Schill // IEEE 5th Int'l Conf. CLOUD. – 2012. – pp. 772-778.
10. Beloglazov A. Energy-Efficient Management of Virtual Machines in Data Centers for Cloud Computing / Beloglazov Anton – Melbourne, 2013.
11. Хританков А.С. Модели и алгоритмы распределения нагрузки. Алгоритмы на основе сетей СМО / А.С. Хританков // Информационные технологии и вычислительные системы. – 2009. – №3. – с. 33–48.
12. Agarwal K. Dynamic Power Management by Combination of Dual Static Supply Voltages / Kanak Agarwal, Kevin Nowka // 8th International Symposium on Quality of Electronic Design (ISQED 2007). – 2007.
13. Gutnik V. Embedded Power-Supply for Low-Power DSP / V. Gutnik, A. Chandrakasan // IEEE Trans on VLSI Systems. – 1997. – Vol. 5, No. 4. – pp. 425-435.
14. Burd T. A Dynamic Voltage Scaled Microprocessor System / T. Burd, T. Pering, A. Stratakos and R. Broderon // IEEE Journal of Solid State Circuits. – 2000. – vol. 35, no. 11. – pp. 1571-1580.
15. Wang S. A DVFS Based Energy-Efficient Tasks Scheduling in a Data Center / S. Wang, Z. Qian, J. Yuan, I. You // IEEE Access. – 2017. – vol. 5. – pp. 13090-13102. doi: 10.1109/ACCESS.2017.2724598
16. Jeabin L. Dynamic Voltage and Frequency Scaling (DVFS) scheme for multi-domains power management / Jeabin Lee, Byeong-Gyu Nam, Hoi-Jun Yoo // 2007 IEEE Asian Solid-State Circuits Conference. – 2007. – pp. 360-363. doi: 10.1109/ASSCC.2007.4425705
17. Aldossary M. Performance and Energy-Based Cost Prediction of Virtual Machines Auto-Scaling in Cloud / Aldossary M, Djemame K. // 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA 2018). – 2018. – pp. 502-509. DOI: 10.1109/SEAA.2018.00086

18. Gandhi A. Optimality analysis of energy-performance trade-off for server farm management / Anshul Gandhi, Varun Gupta , Mor Harchol-Balter, Michael A. Kozuchb // Performance Evaluation. – 2011. – №67. – pp. 1155-1171
19. Möbius C. Power Consumption Estimation Models for Processors, Virtual Machines, and Servers / Christoph Möbius, Waltenegus Dargie, Alexander Schill // IEEE Transactions on Parallel and Distributed Systems. – 2014. – №25. – pp. 1600-1614
20. Gupta V. Optimality analysis of energy-performance trade-off for server farm management / Anshul Gandhi, Varun Gupta , Mor Harchol-Balter, Michael A. Kozuchb // Performance Evaluation. – 2011. – №67. – pp. 1155-1171
21. A.V. Dhumane, R.S. Prasad, J.R. Prasad: An optimal routing algorithm for internet of things enabling technologies, Int. J. Rough Sets Data Anal. (IJRSDA), pp. 1-16, (2017).
22. Pallavi S Katkar and Vijay R Ghorpade : “A Survey on Energy Efficient Routing Protocol for Wireless Sensor Networks”, International Journal of Computer Science and Information Technologies(IJCSIT), vol.6(1), pp.81-83, (2015).
23. Li, Y., Zhang, X., Li, Y.: Algorithm of cluster head multi-hops based on LEACH (September 2007).
24. Banarjee, A. Efficient, Adaptive and Scalable Device Activation for M2M Communications. In Proceedings of the IEEE International Conferences on Sensing, Communications and Networking (SENCON), Seattle, WA, USA, 22–25 June 2015; pp. 399–407, (2015).
25. Network Optimization in the Internet of Things / [Електронний ресурс]– Режим доступу:
<https://www.sciencedirect.com/science/article/pii/S2215098618303379>
26. Проектування інформаційних систем. Навчальний посібник з курсу "Проектування інформаційних систем" [Електронний ресурс] – Режим доступу:
<https://studall.org/all2-114831.html>

27. Методи і засоби підвищення показників якості систем управління телекомунікаційними мережами – Л.Н. Беркман
28. Аналіз вимог і визначення специфікацій програмного забезпечення при об'єктному підході [Електронний ресурс] – Режим доступу: <http://um.co.ua/5/5-3/5-3943.html>
29. Tony Bourke – Server Load Balancing – "O'Reilly Media, Inc.", 2001 – 175 с
30. 1. Балансировка нагрузки: основные алгоритмы и методы – Режим доступу: <HTTPS://habr.com/ru/company/selectel/blog/250201/>. Дата доступу: 10.05.2019